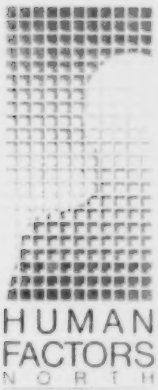
DRDC Toronto CR 2003-046

A COMPLEXITY MEASURE FOR ELECTRONIC DISPLAYS: FINAL REPORT ON THE EXPERIMENTS

by
Don C. Donderi, PhD
Human Factors North Inc.
118 Baldwin Street
Toronto, Ontario
M5T 1L6



Project Manager - D.C. Donderi, PhD
(514) 398-6130
Submitted in partial fulfillment of:
PWGSC Contract No. W7711-017758/001/TOR

On behalf of
DEPARTMENT OF NATIONAL DEFENCE
As represented by:
Defence Research and Development Canada (DRDC)
1133 Sheppard Avenue West
Toronto, Ontario, Canada
M3M 3B9

DRDC Scientific Authority
Sharon McFadden
Phone: (416) 635-2189
March 4, 2003

The scientific or technical validity of this Contract Report is entirely the responsibility of the contractor and the contents do not necessarily have the approval or endorsemenbf Defence R&D Canada.

© HER MAJESTY THE QUEEN AS REPRESENTED BY THE MINISTER OF NATIONAL DEFENCE, 2003
© SA MAJESTÉ LA REINE, REPRÉSENTÉE PAR LE MINISTRE DE LA DÉFENSE NATIONAL, 2003

This study, approved by the McGill University Human Research Ethics Committee, was conducted in conformity with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.

104-06920/

A Complexity Measure for Electronic Displays: Final Report on the Experiments[1]

Mesure de la Complexité des Affichages Électronique: Rapport Définitif sur les Expériences

General Abstract

Five experiments studied the relationship between compressed file size and search performance for visual displays. Experiments 1 and 2 were visual search tasks where observers searched displays to find either a triangle or a parallelogram randomly positioned on the display. In experiment 1 the displays were marine electronic charts, radars, or chart-radar overlays. In experiment 2 they were colour photographs of the earth. In both experiments, search decision time, rated complexity, and proportion of decision errors were monotonically related to the log JPEG compressed file size (LJPG) of the displays. Experiments 3 looked at two different same-different decision tasks in which a marine chart display was followed by a marine radar display. The task was to decide whether they represented the same or a different area. The *ad lib* condition gave unlimited viewing time while the fast condition gave a short viewing time for each display. Chart viewing time in the *ad lib* condition was a positive increasing function of chart LJPG. In the fast condition, decision accuracy depended on chart and radar LJPG and the subjective similarity between chart and radar pairs. Displays in the search tasks of experiments 4 and 5 presented groups of chart symbols containing either triangles or parallelograms as targets and other symbols as background. Decision time in both experiments was non-linear in LJPG, but linear in structural and semantic information content (Experiment 4) or semantic information content (experiment 5). The uses and limitations of compressed file size as a measure of display complexity are discussed.

Résumé Général

Cinq expériences ont été menées en vue d'étudier la relation entre la taille des fichiers compressés et la performance de recherche dans des affichages. Dans les expériences 1 et 2, les observateurs devaient exécuter des tâches de recherche visuelle qui consistaient à examiner des affichages pour trouver un triangle ou un parallélogramme positionné de façon aléatoire dans l'affichage. Les affichages utilisés dans l'expérience 1 étaient composés de cartes électroniques marines, de cartes radar ou de recouvrements carte-radar. Dans l'expérience 2, des photographies couleur de la Terre ont été utilisées. Dans les deux expériences, la durée de la recherche, la cote de complexité et la proportion des erreurs de décision étaient monotoniquement reliées à la taille des fichiers JPEG compressés (TFJC) consignée pour les affichages. L'expérience 3 comprenait deux tâches différentes de distinction qui consistaient à déterminer si une carte marine et une carte radar marine affichées consécutivement représentaient le même emplacement ou un emplacement différent. Dans l'expérience réalisée en condition libre, la durée d'examen des affichages n'étaient pas limitée, tandis que dans l'expérience réalisée en condition

[1] I thank Josie Caro for her diligent and meticulous help in carrying out all of the research described here.

rapide, les observateurs ne disposaient que d'un court laps de temps pour examiner chaque affichage. En condition libre, la durée d'examen de l'affichage était une fonction positive croissante de la TFJC de la carte. En condition rapide, l'exactitude de la décision dépendait de la TFJC de la carte marine et de la carte radar ainsi que de la similarité subjective entre les paires carte marine-carte radar. Pour les tâches de recherche des expériences 4 et 5, les affichages montraient des groupes de symboles de carte dans lesquels des triangles ou des parallélogrammes étaient utilisés comme symboles cibles et d'autres symboles étaient utilisés comme symboles de fond. Le délai de décision (DD) obtenu dans les deux expériences n'était pas une fonction linéaire de la TFJC, mais une fonction linéaire du contenu de l'information structurelle et de l'information sémantique (expérience 4) ou du contenu de l'information sémantique (expérience 5). Le présent document traite de l'utilisation et des limites de la taille des fichiers compressés comme mesure de complexité des affichages.

EXECUTIVE SUMMARY

Five experiments were carried out to determine the relationship between compressed file length (as a measure of visual display complexity) and visual search performance. In experiment 1, forty-three observers searched for a target in marine electronic chart, radar, or chart-radar overlay displays of fifty-three different chart areas. They detected one of two targets, pressed a response key which turned off the display, and then entered a confidence rating. Each display was presented once to each observer in random order. The display type (chart, radar or overlay) shown for each area was counterbalanced across observers. Log decision time (LDT) and rated display complexity both increased with increased log JPEG compressed file size (LJPG), and errors accumulated more rapidly to displays with larger LJPG and higher rated complexity.

In experiment 2, twelve new observers searched for one of two targets in fifty-three colour photo displays of the earth's surface. The displays were presented until the observer detected a target and pressed a response key, after which a confidence rating was obtained. Each display was presented once to each observer in random order. Another group of eleven observers rated the complexity of each display. Search decision time, response errors and rated complexity increased with increased LJPG.

Two same-different tasks were studied in experiment 3. A marine electronic chart display was presented first and followed by a marine radar display. The task was to decide whether the geographical areas of the two displays were the same or different and then to make a confidence judgment. In the *ad lib* condition, display duration was at the observer's discretion for both the chart and the radar display. In the *fast* condition, each display was presented for 0.25 sec with an 0.1 sec inter-display interval. Chart viewing time (VT) in the ad lib experiment was a positive increasing function of chart LJPG. Decision accuracy increased with longer radar VT, larger radar LJPG and it was related to the subjective similarity of chart and radar pairs. Accuracy was higher for same display pairs with higher similarity ratings, and higher for different display pairs with lower similarity ratings. Decision accuracy in the fast experiment depended on both chart and radar LJPG and the subjective similarity between chart and radar pairs.

A new type of visual display was studied in experiments 4 and 5. In experiment 4, observers searched 48 displays containing 36 chart symbols to find one, two or three target symbols. The target symbols were either triangles or parallelograms; the background symbols were other simple chart symbol shapes. The displays differed in the number of different symbol types displayed and in the number and arrangement of symbol groups on the display. Decision latency was an increasing function of the number of different symbols displayed, the number of symbol groups on the display, and the constraints placed on symbol group positions across the display. The results show that a structural analysis of alphanumeric display complexity can be applied to non-alphanumeric symbolic displays.

In experiment 5, eleven observers searched a set of 72 displays composed of 24 chart symbols to find either a triangle or parallelogram target among the remaining different background symbols. There were either 1, 2 or 3 different background symbols arranged in six groups of four, twelve groups of two or individually according to one of two rules: either the x and y coordinates for each symbol (or group) were independent, or the groups were constrained to fall in rows and columns. Changing the number of symbols varied the semantic information while changing the number of groups and the placement constraint varied the structural information. Log decision time (LDT) increased with an increase in rated complexity and semantic information content. Another group of twelve observers rated each display for complexity, which increased with an increase in semantic and structural information.

The experiments show that JPEG file size, an objective measure that is easily obtained for computer-based displays, correlates with performance measures obtained from target search and same-different decision tasks carried out with marine electronic chart, radar and overlay displays and with aerial colour photographs. These results complement the earlier discovery that JPEG file size was correlated with subjective complexity judgments obtained for the same sets of displays.

The non – monotonic effect of LJPG on LDT and complexity ratings found in experiments 4 and 5 limits the generality of the compressed file size measure of visual display complexity. Instead, these experiments suggest that measures of semantic or structural information content, independent of compressed file size, predict performance on these chart symbol displays. But for at least some displays, JPEG file size is a simple, practical and objective display measure that correlates with the subjective experience of display complexity and with objective measures of display performance.

SOMMAIRE

Cinq expériences ont été menées pour déterminer la relation entre la taille des fichiers compressés (comme mesure de la complexité d'un affichage) et la performance de recherche visuelle. Dans l'expérience 1, quarante-trois observateurs devaient trouver, dans cinquante-trois affichages d'emplacements différents, une cible présentée sur une carte électronique marine, sur une carte radar ou sur un recouvrement carte-radar. Ils devaient détecter une des deux cibles, appuyer sur une touche de réponse pour supprimer l'affichage, puis entrer un taux de confiance. Chaque affichage a été présenté, en ordre

aléatoire, une fois à chacun des observateurs. Les types d'affichage (carte marine, carte radar ou recouvrement) utilisé pour chaque emplacement ont été répartis également entre les observateurs. Le délai de décision (DD) consigné et la complexité cotée de l'affichage ont tous les deux augmenté en même temps que la taille du fichier JPEG compressé (TFJC) consignée, et les erreurs s'accumulaient plus rapidement pour les affichages dont la TFJC et la complexité cotée étaient plus élevées.

Dans l'expérience 2, douze nouveaux observateurs devaient trouver une des deux cibles présentées dans cinquante-trois affichages de photographies couleur de la surface de la Terre. Les affichages étaient montrés jusqu'à ce que l'observateur détecte une cible et appuie sur une touche de réponse, puis entre un taux de confiance. Chaque affichage a été présenté, en ordre aléatoire, une fois à chacun des observateurs. Un autre groupe de onze observateurs ont coté la complexité de chaque affichage. Le délai de décision, le nombre d'erreurs et la complexité cotée augmentaient lorsque la TFJC augmentait.

Deux tâches de distinction ont été étudiées dans le cadre de l'expérience 3. Un affichage de carte électronique marine a d'abord été présenté, puis un affichage de carte radar marine. La tâche consistait à déterminer si l'emplacement géographique montré dans les deux affichages était le même ou différent, puis à indiquer le niveau de confiance de la réponse. En condition libre, l'observateur était libre d'examiner l'affichage de la carte marine et de la carte radar aussi longtemps qu'il le voulait. En condition rapide, chaque affichage était présenté pendant 0,25 s avec une intervalle de 0,1 s entre chaque affichage. Dans l'expérience menée en condition libre, la durée d'examen de la carte (DEC) était une fonction positive croissante de la TFJC de la carte. L'exactitude des décisions était meilleure lorsque la DEC radar et la TFJC de la carte radar étaient plus élevées, et était liée à la similarité subjective des paires carte marine-carte radar. L'exactitude était plus élevée pour les paires d'affichages pareils ayant une cote de similarité plus élevée, et également pour les paires d'affichages différents ayant une cote de similarité plus basse. Dans l'expérience menée en condition rapide, l'exactitude des décisions dépendait de la TFJC des cartes marines et des cartes radar et de la similarité subjective des paires carte marine-carte radar.

Un nouveau type d'affichage a été étudié dans les expériences 4 et 5. Dans l'expérience 4, les observateurs devaient examiner 48 affichages contenant 36 symboles de carte pour trouver un, deux ou trois symboles cibles. Les symboles cibles étaient soit des triangles, soit des parallélogrammes. Les symboles de fond étaient des symboles de carte ayant une forme simple. Le nombre de types de symboles différents, ainsi que le nombre et la disposition des groupes de symboles variaient d'un affichage à l'autre. Le délai de décision était une fonction croissante du nombre de symboles différents affichés, du nombre de groupes de symboles dans chaque affichage et des contraintes imposées sur la position des groupes de symboles dans l'affichage. Les résultats montrent que l'analyse structurelle de la complexité des affichages alphanumériques peut s'appliquer aux affichages de symboles non alphanumériques.

Dans l'expérience 5, onze observateurs ont examiné un ensemble de 72 affichages composés de 24 symboles de carte pour trouver une cible représentée soit par un triangle,

soit par un parallélogramme, parmi les différents symboles de fond. Un, deux ou trois symboles différents étaient disposés en six groupes de quatre, douze groupes de deux, ou individuellement, conformément à l'une des deux règles suivantes : les coordonnées x et y de chaque symbole (ou groupe de symboles) étaient indépendantes les unes des autres ou les groupes devaient former des rangées et des colonnes. Une variation du nombre de symboles entraînait une modification de l'information sémantique, tandis qu'une variation du nombre de groupes et de la position des symboles entraînait une modification de l'information structurelle. Le délai de décision (DD) consigné augmentait lorsque la complexité cotée et le contenu de l'information sémantique augmentaient. Un autre groupe de douze observateurs ont coté la complexité de chaque affichage, qui était plus élevée lorsque la quantité d'information sémantique et d'information structurelle augmentait.

Les expériences ont démontré qu'une corrélation existe entre la taille des fichiers JPEG, une mesure objective facile à obtenir pour les affichages sur ordinateur, et les mesures de performance obtenues lors de la recherche de cible et de l'exécution des tâches de distinction de symboles figurant sur des cartes électroniques marines, sur des cartes radar, sur des recouvrements carte-radar et sur des photographies couleur. Ces résultats complètent la recherche précédente qui a permis d'établir une corrélation entre la taille des fichiers JPEG et les évaluations subjectives de la complexité du même ensemble d'affichages.

En raison de l'effet non monotonique de la TFJC sur le DD et des cotes de complexité constatées dans les expériences 4 et 5, on ne peut pas généraliser l'utilisation de la taille des fichiers compressés comme mesure de complexité des affichages. Ces expériences suggèrent plutôt que la mesure du contenu de l'information sémantique ou de l'information structurelle, indépendamment de la taille des fichiers compressés, permet de prévoir la performance d'observation des affichages de symboles de carte. Cependant, pour certains types d'affichages, la taille des fichiers JPEG constitue une mesure simple, pratique et objective des affichages, qui est en corrélation avec la complexité subjective des affichages et avec les mesures objectives de la performance d'observation des affichages.

TABLE OF CONTENTS

GENERAL INTRODUCTION 1

EXPERIMENT 1 3

ABSTRACT 3
INTRODUCTION 3
RESULTS 4
DISCUSSION 7

EXPERIMENT 2: 11

ABSTRACT 11
INTRODUCTION 11
INTRODUCTION 11
METHOD 11
Search task *11*
Complexity task *13*
RESULTS 13
DISCUSSION 13

EXPERIMENT 3 16

ABSTRACT 16
INTRODUCTION 16
METHOD 16
Displays *17*
Procedure *17*
RESULTS 17
Ad Lib condition *17*
Fast condition *21*
DISCUSSION 23

EXPERIMENT 4 24

ABSTRACT 24
INTRODUCTION 24
METHOD 26
Stimuli *27*
Display information measures *30*
Instructions *30*
Presentation conditions *30*
RESULTS 30
Information measures *33*
DISCUSSION 35

EXPERIMENT 5 36

ABSTRACT 36
INTRODUCTION 36
METHOD 36
Displays *36*
Presentation *37*

Complexity Rating 37
RESULTS 38

SUMMARY AND CONCLUSIONS 45

EXPERIMENTAL FINDINGS 45
Experiment 1: 45
Experiment 2 45
Experiments 4 and 5 46
CONCLUSIONS 46
Search Experiments and JPEG File Size 46
Same – Different Experiments 47
Chart Symbol Search Experiments 48
JPEG File Size and Complexity 48
JPEG Complexity, Algorithmic Complexity and Random Complexity 49
Unanswered questions 50

REFERENCES 53

List of Figures

Figure 1. Experiment 1: Chart (top), radar (middle) and overlay (bottom) displays 5
Figure 2. Experiment 1: JPEG file size as a function of ZIP file size for radar (R), chart (C) and overlay (O) displays. 8
Figure 3. Experiment 1: Log decision time as a function of log JPEG file size. 8
Figure 4. Experiment 1: Mean log decision time as a function of log JPEG file size averaged over displays. 9
Figure 5. Experiment 1: Mean complexity rating as a function of log JPEG file size averaged over displays. 9
Figure 6. Experiment 1: Mean log decision time per display as a function of mean complexity rating. 10
Figure 7. Experiment 1: Summary of maximum decision time, maximum complexity and cumulative error as a function of log JPEG file size. 10
Figure 8. Experiment 2: Two photos similar to those used as displays. 12
Figure 9. Experiment 2: Log JPEG file size as a function of log ZIP file size. 14
Figure 10. Experiment 2: Log decision time as a function of log JPEG file size for all responses. 14
Figure 11. Experiment 2: Mean complexity as a function of log JPEG file averaged over displays. 15
Figure 12. Experiment 2: Summary of maximum decision time, maximum complexity and cumulative error as a function of log JPEG file size. 15
Figure 13. Experiment 3: Log chart viewing time as a function of log chart JPEG file size. 19
Figure 14. Experiment 3: Mean error proportion as a function of log radar JPEG file size. 19
Figure 15. Experiment 3, ad lib condition: Effect of dissimilarity on proportion correct and radar viewing time for same and different pairs of experiment 3. 20
Figure 16. Experiment 3, fast condition: proportion correct as a function of log chart JPEG file size. 20
Figure 17. Experiment 3, fast condition: Proportion correct as a function of radar log JPEG file size. 22
Figure 18. Experiment 3, fast condition: Proportion correct as a function of mean chart-radar dissimilarity. 22
Figure 19. Experiment 4: Alphanumeric screen (above) with symbol group rectangles outlined (below). From Tullis (1983). 25
Figure 20. Two alphanumeric displays with identical structural content and different semantic content. 26
Figure 21. Experiment 4: Structurally and semantically simple (top) and complex (bottom) chart symbol displays 29
Figure 22. Experiment 4: Mean log decision time as a function of the number of groups and the number of symbols. 32

FIGURE 23. EXPERIMENT 4: MEAN LOG DECISION TIME AS A FUNCTION OF THE SUM OF SEMANTIC AND STRUCTURAL INFORMATION FOR ALL RESPONSES..... 32
FIGURE 24. EXPERIMENT 4: MEAN LOG DECISION TIME AS A FUNCTION OF LOG JPEG FILE SIZE 34
FIGURE 25. EXPERIMENT 5: LOG JPEG FILE SIZE AS A FUNCTION OF LOG ZIP FILE SIZE FOR THE CHART SYMBOL FIGURES. 34
FIGURE 26. EXPERIMENT 5: SIMPLE (TOP) AND COMPLEX (BOTTOM)) CHART SYMBOL DISPLAY 39
FIGURE 27. EXPERIMENT 5: MEAN LOG DECISION TIME AND MEAN PROPORTION OF ERRORS AS A FUNCTION OF SEMANTIC INFORMATION IN BITS. 42
FIGURE 28. EXPERIMENT 5: MEAN COMPLEXITY AS A FUNCTION OF SEMANTIC AND STRUCTURAL INFORMATION. 42
FIGURE 29. EXPERIMENT 5: PREDICTION OF COMPLEXITY RATING FOR ALL DISPLAYS FROM SEMANTIC AND STRUCTURAL DISPLAY INFORMATION. 43
FIGURE 30. EXPERIMENT 5: MEAN LOG DECISION TIME AS A FUNCTION OF MEAN LOG COMPLEXITY RATING. 43
FIGURE 31. EXPERIMENT 5: MEAN LOG DECISION LATENCY AS A FUNCTION OF MEAN LOG JPEG FILE SIZE..... 44
FIGURE 32. EXPERIMENT 5: MEAN COMPLEXITY RATING AS A FUNCTION OF MEAN LOG JPEG FILE SIZE..... 44

List of Tables

Table 1. Experiment 1: Mean and standard deviation of correct decision time (.01 sec) and number and percent of correct decisions for display types, display sets and observer groups* 6

Table 2. Experiment 1: Mean and SD of correct decision time (.01 sec), N (Number of Cases) and Percent of Right and Wrong Decisions for Each Display Type ... 6

Table 3. Experiment 1: Analysis of covariance of Log Decision Time and Correct Responses. 7

Table 4. Correlations among Variables across Display Pairs for the Ad Lib Experiment 18

Table 5. Partial correlation coefficients predicting mean accuracy from JPEG and latency variables. 18

Table 6. Fast Same-Different Experiment: Standardized Regression Coefficients and Partial Correlation Coefficients relating Probability of Correct Response to Each Independent Variable 21

Table 7. Information analysis of the Displays used in Experiment 4. 28

Table 8. Mean and SD of Decision Latency, and Proportion of Correct Responses, for Number of Background Symbols, Number of Groups, and Random or Constrained Group Coordinates in the Symbol Search Experiment 31

Table 9. Semantic and Structural Information Measures for the displays in Experiment 5. 37

Table 10. Log decision time and (SD) for Display Design Factors: Symbols, Groups and Rules. 40

GENERAL INTRODUCTION

This report describes five experiments on the relationship between visual display complexity and visual search performance. Previous research showed that the compressed file size of marine electronic chart, marine radar and chart-radar overlay displays predicted the subjective complexity of the displays (Donderi 2001, 2000,1999; Donderi & McFadden, 1999). The goal of the experiments reported here was to learn whether compressed file size was related to objective performance with visual displays.

There is reason to think that compressed file size might be related to subjective complexity and to objective performance. A compressed computer file is analogous to one part of a very general description of complexity called algorithmic probability. Solomonoff (1986) defines algorithmic probability as follows:

> Algorithmic probability of a string, s, is the probability of that particular string being produced as the output of a reference universal Turing machine with random input. It is approximately $2^{-|p|}$ where p is the length of the shortest program, p, that will produce s as an output. p is the Kolmogorov complexity of s: one form of algorithmic complexity.

Since algorithmic probability (or Kolmogorov complexity) provides a metric for the complexity of any symbol string, it can be used to measure the complexity of the bit string that constitutes a visual display image. The algorithmic probability computation has three parts: 1) the computing algorithm (the Turing machine), 2) the shortest program p that drives the algorithm to produce the string, and 3) the result (the string s whose complexity is to be measured). It is obviously impractical to measure the complexity of a visual display by setting up a universal Turing machine and then writing many programs for it in order to find the shortest program that can generate the display. However there is a simple modern analogue to algorithmic probability. Taking the conceptual framework for algorithmic probability one part at a time 1) The computing algorithm that substitutes for the Turing machine is a file compression program, 2) The shortest program, p that generates the visual display is the compressed image file, and 3) its result, s is the string of pixels that reconstitutes the visual display. In this analogy, the compressed file is the generating program whose length p is the complexity measure of the visual display. The file compression program acts as an inverse Turing machine when it creates a shortest program from the image, and as a Turing machine when it re-creates the image from the shortest program.

Leeuwenberg (1968, 1969) developed a code that specified the information in any patterns that could be described as the concatenation of a series of repetitive elements. The information measure based on his code accurately predicted both complexity measures and performance measures for some displays, but he said in 1969 "for the time being [this coding procedure] will hardly lend itself to computer programming." – which cannot be said today about JPEG or ZIP compression coding.

The algorithmic probability analogy can be put to the test. Start with a set of visual displays, obtain subjective complexity and objective performance measures, pass the displays through the WinZip and JPEG compression programs (used, in all the experiments, in the lossless mode), count the number of bytes in the compressed files, and compare this measure of complexity with the subjective complexity and the measured performance. It is easier to start with electronic displays, but any visual display of reasonable size can be scanned into a computer and its bitmap file then compressed.

The search task studied in experiment 1 and the same-different decision tasks studied in experiment 3 were carried out with the chart, radar and overlay displays used in the earlier research. All three experiments showed strong relationships between JPEG file size, decision time and decision accuracy. The target search task studied in Experiment 2 was based on aerial colour photographs and the results replicated Experiment 1. Experiments 4 and 5 used chart symbol displays in order to investigate Tullis' (1983) distinction between structural and semantic measures of display complexity. Both experiments showed strong relationship between semantic complexity and search performance and an inverted U- relationship between search performance and compressed file size. Experiment 5 also showed a strong linear relationship between judged complexity and measured semantic and structural complexity.

The report is organized into five sections, each dealing with a single experiment. A sixth section summarizes the experimental results followed by a discussion and conclusions. All of the experiments were reviewed and approved by the McGill University Research Ethics Committee for human experiments.

EXPERIMENT 1: TARGET SEARCH DECISION TIME AND ACCURACY WITH ELECTRONIC CHART, OVERLAY AND RADAR DISPLAYS ARE PREDICTED BY JPEG FILE SIZE

Abstract

Forty-three observers searched for a target in marine electronic chart, radar, or chart-radar overlay displays of fifty-three different chart areas. They detected one of two targets, pressed a response key which turned off the display, and then entered a confidence rating. Each display was presented once to each observer in random order. The display type (chart, radar or overlay) shown for each area was counterbalanced across observers. Log decision time (LDT) increased with increased log JPEG compressed file size (LJPG) and subjective display complexity, and errors accumulated more rapidly to displays with larger LJPG and higher rated complexity.

Introduction

The goal of a visual search experiment is to find the relationship between properties of a visual display and the duration and accuracy of a search for one or more targets in the display. If the search target is randomly positioned, then, on the average, over many displays, the display property under investigation should influence the average decision time to identify the target. The property under study in this experiment was the log JPEG file size of the display (LJPG), which is already known to predict rated display complexity (Donderi, 2000). The goal was to learn whether LJPG also predicted search performance.

Method

Each display contained one of two targets, a small triangle or a small parallelogram, randomly selected and randomly positioned on the display. The task was to decide which target was present and then to make a confidence judgment about the decision. The display was presented until a decision was made. Instructions were to respond as rapidly as possible consistent with accuracy. Decision time in .01 sec units, accuracy, and confidence (on a scale from 1 to 9, with 1 indicating guessing and 9 certainty) were recorded after each display presentation.

The displays were either marine electronic charts, marine radar images or overlays created from marine charts combined with the aligned and scaled radar images of the same areas displayed transparently over the charts. The entire set of displays had been used in an earlier study that evaluated the efficiency of overlaid versus separate presentation of marine electronic charts and radar (Donderi, 1999). Complexity and dissimilarity ratings had also been obtained for the displays (Donderi, 2000). There were 53 distinct display areas, each represented once by each type of display. The areas were based on harbour charts of Halifax, N.S., St. John's, Nfld., New York, N.Y., and Hammarsfjord, Norway. Three display groups of 53 displays each were created so that each area was represented once within each display group but, within that group, in only

one of the three display types (chart, radar or overlay). The displays in each group represented the four harbour areas about equally often. One-third (17 or 18) of the 53 displays seen by each observer were charts, one-third were radars and one-third were overlays. Each of the 43 observers saw one of the display groups, so each observer saw a total of 53 mixed chart, radar and overlay displays. The displays in each group were unique and no observer saw the same display or the same area twice. Sample displays are shown in Figure 1.

The target location was determined by random selection of the x and y coordinates on each display, and which target appeared was also determined by random selection for each display. The targets were triangle and parallelogram chart symbols from a set studied by McFadden (personal communication). The displayed target dimensions were 2 mm wide by 2 mm high, subtending about 0.4 deg when viewed at 30 cm. Each display was 21 cm wide by 21 cm high, subtending an angle of about 35 deg at 30 cm. The targets were presented transparently over the display image.

The entire experiment was presented on a 17 in flat-screen CRT monitor. At the beginning of the experiment, observers read on-screen instructions to search each display for as long as necessary to determine which target (triangle or parallelogram) was present, and then to press one of two keys (the bracket keys) labelled with a triangle and parallelogram respectively. Display groups were alternated among observers and the target key symbols were alternated between observers. Instructions were to respond as rapidly as possible, consistent with accuracy. The displays were presented until the observer made a decision. When the decision key was pressed, the display disappeared. Decision time to the nearest hundredth of a second was recorded from display onset to the decision key press. Then observers made a decision confidence rating from 1 to 9, with 1 meaning not at all sure and 9 meaning absolutely certain. After entering the confidence rating in a box on the screen, the observer was instructed to press "enter" when ready for the next display. The next display followed two seconds later. Participants started with a practice set containing one chart, one radar and one overlay display that did not appear among the 53 experimental displays. Practice data was discarded. Then the sequence of 53 experimental displays was presented to each observer in an independently randomized order. The entire experiment took about 25 minutes per observer.

Results

ZIP and JPEG compressed file sizes were highly correlated over the display image set. JPEG sizes were smaller than ZIP sizes for all displays, but Figure 2 shows that JPEG compression was more efficient for chart and overlay than for radar displays. Salomon (1997, p. 165) observed that JPEG "does not handle bi-level (black and white) images very well."

Log decision times (LDT) for all responses are plotted as a function of log JPEG display file size in Figure 3. The linear regression of LDT on LJPG is shown by the black line.

Figure 1. **Experiment 1: Chart (top), radar (middle) and overlay (bottom) displays**

The envelope of maximum LDT up to each JPEG value, as well as the linear regression based on the maxima, are shown in blue. Cumulative errors are plotted in red. The error slope is almost flat until LJPG exceeds about 4.9, after which it rises steeply. Figure 4 shows the same data averaged over each display. It shows that LDTs for radar displays were faster than LDTs for chart and overlay displays and that most of the errors occurred to the overlay displays with large JPEG file sizes.

Tables 1 and 2 summarize the results that decisions were fastest and most accurate to radar displays and slowest and least accurate to the overlay displays and that wrong decisions took longer and were more variable than correct decisions for all display types. Analysis of covariance of LDTs and correct responses with observers within groups as the random effect and order of presentation as a covariate showed (Table 3) that both LDT and the number of correct responses differed significantly across display types. Display type differences for LDT and correct responses were significant at $p < 0.05$ (Scheffe multiple-comparison test).

Table 1. Experiment 1: Mean and standard deviation of correct decision time (.01 sec) and number and percent of correct decisions for display types, display sets and observer groups*

Observer Group	Display type											
	Charts				Overlays				Radars			
	Set	Mean	SD	N (%)	Set	Mean	SD	N (%)	Set	Mean	SD	N (%)
1 (O = 15)	A (D=18)	230	229	253 (94)	C	334	626	264 (98)	B	215	197	254 (100)
2 (O = 14)	B (D=17)	287	365	207 (87)	A	304	532	133 (53)	C	171	95	231 (92)
3 (O = 14)	C (D=17)	457	584	215 (90)	B	819	1212	138 (55)	A	231	480	228 (90)
Average		320	422	675 (90)		451	829	535 (69)		206	301	713 (94)

*The maximum possible number (N) of correct responses differed across groups and types because the number of observers and the number of displays were not equal across groups and types. O = number of observers per group, D = number of displays per set..

Table 2. Experiment 1: Mean and SD of correct decision time (.01 sec), N (Number of Cases) and Percent of Right and Wrong Decisions for Each Display Type

Display type	Decision	Mean	SD	N	(%)
Charts	Right	320	422	675	(95)
	Wrong	691	761	37	(05)
Overlays	Right	451	829	535	(72)
	Wrong	576	849	203	(28)
Radars	Right	206	301	713	(99)
	Wrong	345	650	10	(01)

Table 3. Experiment 1: Analysis of covariance of Log Decision Time and Correct Responses.

Analysis factor	Log decision time		Correct responses	
	MS	F	MS	F
Display type	6.25	134.73*	16.25	697.76*
Display set	1.07	23.05*	5.23	224.58*
Type x set	0.55	11.90*	3.65	156.92*
Observers within groups	0.75	16.10*	0.19	0.80
Displays within type x set	0.89	19.08*	0.80	34.15*
Presentation order	2.41	51.86*	0.03	1.32
Error	0.05		0.02	

* $p < 0.0001$

Figure 5 plots the mean complexity rating for each display (obtained in earlier studies: Donderi, 1999, 2000) against the LJPG for that display. Like LDT, mean complexity increases as a function of LJPG. Figure 6 plots the mean LDT for each display against the mean display complexity rating. Comparison of Figures 5 and 3 shows that rated complexity, like LDT, increases with increased LJPG. The results of the experiment are summarized in Figure 7 which plots the relationship between LJPG and maximum LDT, rated complexity and cumulative error probability.

Discussion

Figure 3 shows that there are frequent fast and accurate decisions to displays with large LJPG, but that the average and the maximum decision time both increase as LJPG increases. The plot of cumulative errors against LJPG is not linear, as it would be if errors accumulated at the same rate regardless of LJPG. Instead, there is a threshold at about 4.9 LJPG before which decision errors are rare and beyond which errors accumulate rapidly as LJPG increases.

Radar displays with the smallest average LJPG had the fastest decision times and the largest percentage of correct decisions. The chart displays had the next larger average LJPGs and were both slower and less accurate, and the overlay displays with the largest LJPGs had the slowest and least accurate responses.

In earlier research, the rated complexity of the chart, radar and overlay displays was shown to increase with increasing LJPG. In this experiment, LDT and errors increased as both rated complexity and LJPG increased. This suggests that LJPG measures the subjective complexity of a visual display, and that it predicts speed and accuracy in search performance. Even though decision time and LJPG were logarithmically transformed, both the maximum and the dispersion of LDT increased as LJPG increased. The search process in these marine electronic displays is highly sensitive to an increase in both rated display complexity and its physical analogue, LJPG.

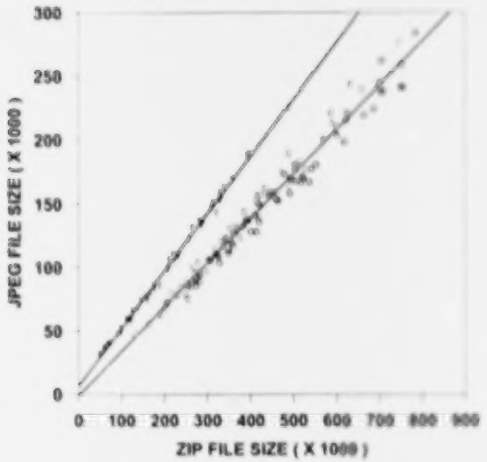


Figure 2. **Experiment 1: JPEG file size as a function of ZIP file size for radar (r), chart (c) and overlay (o) displays.**

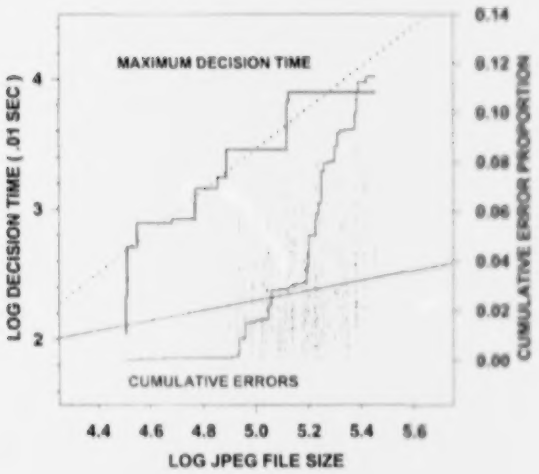


Figure 3. **Experiment 1: Log decision time as a function of log JPEG file size.**

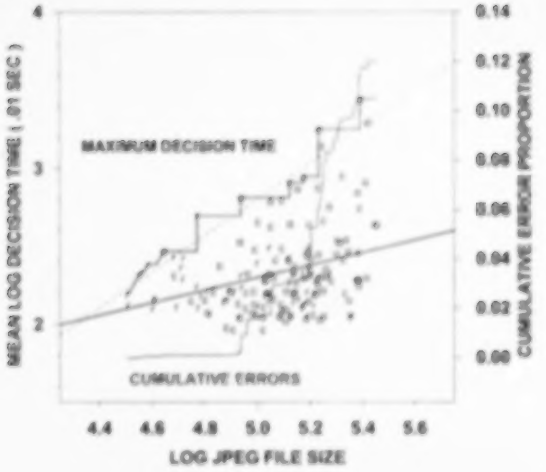


Figure 4. **Experiment 1: Mean log decision time as a function of log JPEG file size averaged over displays.**

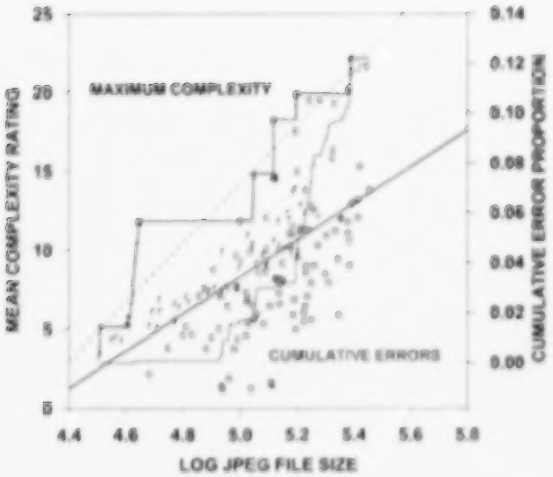


Figure 5. **Experiment 1: Mean complexity rating as a function of log JPEG file size averaged over displays.**

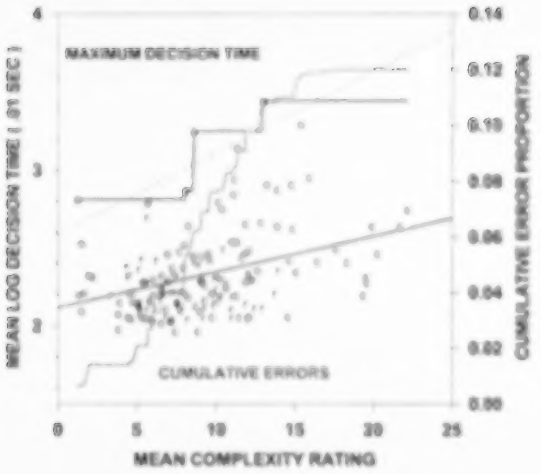


Figure 6. **Experiment 1: Mean log decision time per display as a function of mean complexity rating.**

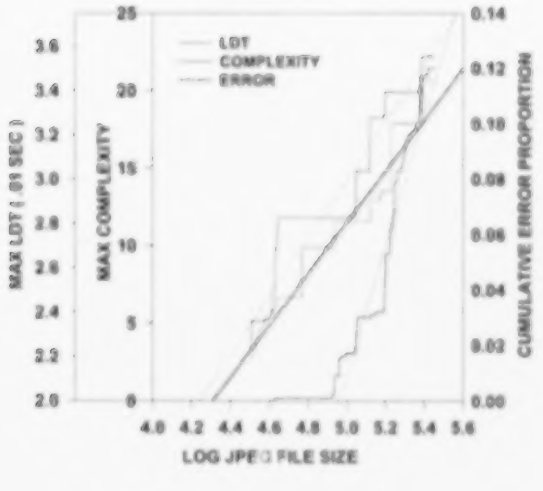


Figure 7. **Experiment 1: Summary of maximum decision time, maximum complexity and cumulative error as a function of log JPEG file size.**

EXPERIMENT 2: VISUAL SEARCH WITH COLOUR PHOTO DISPLAYS

Abstract

Twelve new observers searched for one of two targets in fifty-three colour photo displays of the earth's surface photographed from above. The displays were presented until the observer detected a target and pressed a response key, after which a confidence rating was obtained. Each display was presented once to each observer in random order. Another group of eleven observers rated the complexity of each display. Search decision time, response errors and rated complexity increased with increased LJPG.

Introduction

Earlier research (Donderi, 1999) showed that subjective complexity ratings of a set of twenty images from the book "Earth from Above" (Arthus-Bertrand, 1999) were correlated with the JPEG compressed file size of each image. The goal of this experiment was to try to replicate the original subjective complexity finding with a larger set of images and to study visual search performance using the same images. The experiment thus followed the pattern of Experiment 1 where the goal was to study the relationship among subjective complexity measures, search performance, and compressed file size.

Method

Displays

The twenty previously studied earth images were used plus 33 new ones to generate a display set of 53 images. The displays were an eclectic selection including urban landscapes, ice floes, desert and jungle scenes, rural agriculture and rural dwellings, some with people, and some without. One-half (one page) of each of the selected double-page images from the Arthus-Bertrand book was scanned into an 1.11 megabyte bitmap computer file. Two sample displays are shown in Figure 8.

Search task

The method was the same as in Experiment 1. Each display contained a single triangle or parallelogram target (the same as in Experiment 1), randomly selected and randomly positioned on the display. The task was to decide which target was present and to make a confidence judgment about the decision. The display remained visible until the decision was made. Instructions were to respond as rapidly as possible consistent with accuracy. Decision time, accuracy and confidence on a 1 (guessing) to 9 (certain) scale were recorded. The test displays were preceded by a practice set of 6 different displays, and the data from the practice set data were discarded. Each test display was presented to each observer in a different random order.

Figure 8. **Experiment 2: Two photos similar to those used as displays.**

Complexity task

A set of eleven different observers made complexity ratings for each display. Instructions told the observer to look at each display and to judge its complexity, using any numbers greater than zero, and with larger numbers indicating greater complexity. The instructions were that if one display appeared to be twice as complex as another, it should be given a complexity rating twice as large. The observer was free to choose his or her own range of numbers. The displays were presented in random order and the display presentation ended when the observer pressed a key before entering the numerical rating. Presentation conditions were the same as in the search task.

Results

ZIP and JPEG compressed file sizes were compared for the 53 displays. Figure 9 shows that they are linearly related (on log coordinates), and that ZIP and JPEG file sizes were about equal across this display set.

Figure 10 plots LDT as a function of LJPG for all of the data. The LDT maximum and the cumulative proportion of errors are also plotted. The regression of LDT on LJPG was significant [$F(3,568) = 7.2$, $p < 0.0001$], accounting for three percent of the variance of all LDT responses. Presentation order had no effect on the result, but both linear and quadratic log JPEG terms were significant at $p < 0.01$. Figure 10 shows that errors were infrequent until LJPG reached about 6.3, after which they increased rapidly.

Figure 11 plots the mean complexity rating against LJPG for each display. The linear regression accounted for a significant ($p < 0.0001$) thirty-one percent of the variance. Figure 12 summarizes the experimental results by plotting the maximum LDT, maximum complexity and cumulative proportion of errors as functions of LJPG.

Discussion

LJPG and log ZIP file sizes were again found to be highly correlated for the colour photo displays, confirming the high correlation found with the marine electronic displays. Comparison of Figures 3 and 10 shows the same relationship between LDT and LJPG in both experiments. There are fast and accurate responses at every value of LJPG, but as LJPG increases, so does the maximum LDT. Once LJPG passes a threshold value, the probability of error increases rapidly.

Rated complexity increased with increasing LJPG, replicating the results of earlier studies on the marine electronic displays. Experiments 1 and 2 confirm that LJPG predicts rated complexity, decision time and accuracy in a simple search task for two different types of graphic displays.

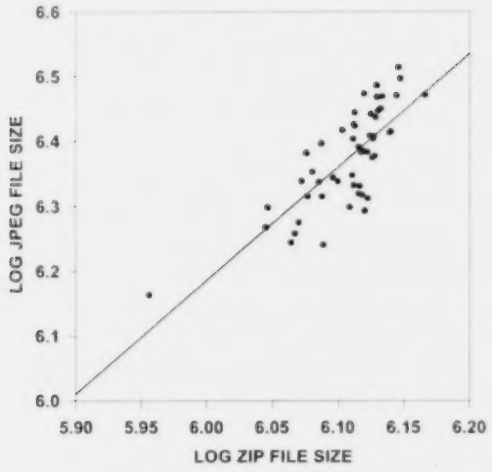


Figure 9. **Experiment 2: Log JPEG file size as a function of log ZIP file size.**

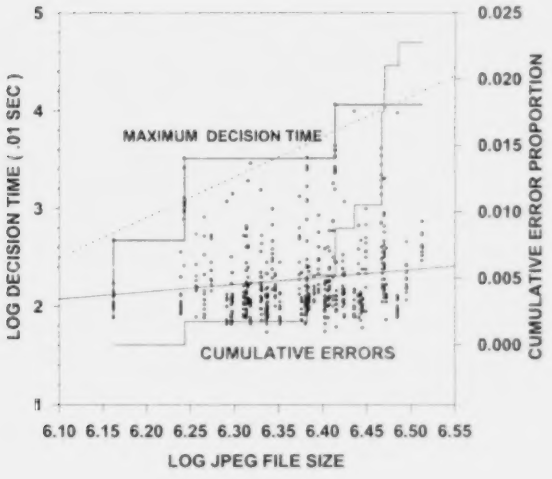


Figure 10. **Experiment 2: Log decision time as a function of log JPEG file size for all responses.**

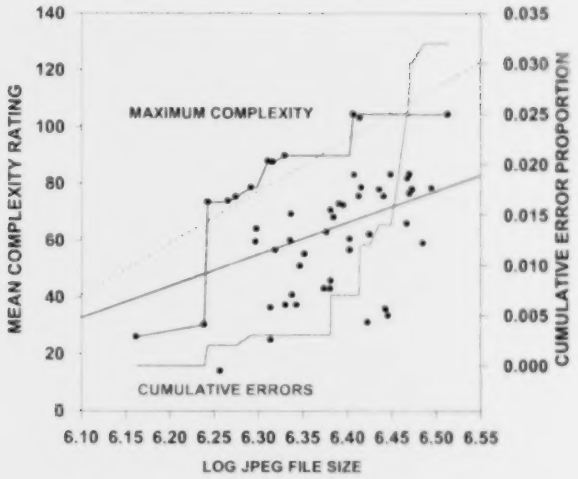


Figure 11. **Experiment 2: Mean complexity as a function of log JPEG file size averaged over displays.**

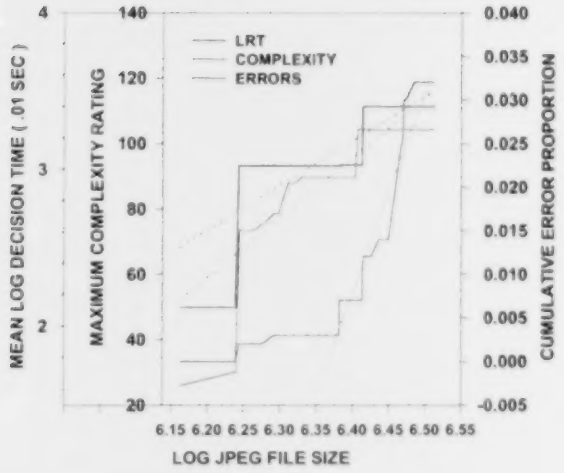


Figure 12. **Experiment 2: Summary of maximum decision time, maximum complexity and cumulative error as a function of log JPEG file size.**

EXPERIMENT 3: THE EFFECT OF JPEG FILE SIZE ON SAME-DIFFERENT MATCHING OF RADAR TO CHART DISPLAYS

Abstract

A marine electronic chart display was presented first and followed by a marine radar display. The task was to decide whether the geographical areas of the two displays were the same or different and then to make a confidence judgment. In the *ad lib* condition, display duration was the observer's choice for both the chart and the radar display. In the *fast* condition, each display was presented for 0.25 sec with an 0.1 sec inter-display interval. Chart viewing time (VT) in the ad lib experiment was a positive increasing function of chart LJPG. Decision accuracy increased with longer radar VT, larger radar LJPG and was related to the subjective similarity of chart and radar pairs. Accuracy was higher for same display pairs with higher similarity ratings and higher for different display pairs with lower similarity ratings. Decision accuracy in the fast experiment depended on both chart and radar LJPG and the subjective similarity between chart and radar pairs.

Introduction

The rated complexity of marine electronic chart, radar and chart-radar overlay displays can be predicted from LJPG (Donderi, 2000). This result led us to study tasks that might also be related to LJPG. One such task is matching a radar to a chart display. A navigator viewing an electronic chart display on one screen and a radar display on another screen must be sure that they represent the same geographical area. A mismatch could happen because of a position error in the navigation system. How long does it take to decide, accurately, that the radar does or does not match the chart? The matching decision can be studied in the framework of a same-different experiment.

Method

Before participating in the same-different experiment each new observer read a description of marine electronic charts and marine radar, and was shown examples of a marine electronic chart and its corresponding radar display. Observers first saw a marine electronic chart display and then saw a marine radar display. Their task was to decide whether the geographic area represented by the marine radar display was the same or different than the area represented by the marine chart display. Each observer participated in either an *ad lib* (N = 21) or a *fast* (N = 22) display condition. In the ad lib condition, the chart display was presented until the observer pressed the space bar, after which the chart display was replaced by the radar display. When the observer pressed a key labelled "S" (same) or "D" (different; the bracket keys on a standard PC keyboard) the radar display disappeared and the decision confidence rating box appeared. The decision and the confidence rating were recorded, and the display latency of both the chart and radar displays was recorded to the nearest hundredth of a second. In the fast condition, the

chart display was presented for 0.25 sec, there was an 0.10 sec delay, and then the radar display was presented for 0.25 sec. The observer then pressed the keys labelled "S" or "D" and entered a decision, followed by a confidence rating.

Displays

The same set of 48 chart-radar display pairs was used for both the ad lib and fast conditions. Half of the pairs consisted of matched chart and radar displays (correct same response) and half of mismatched displays (correct different response). The charts and radar displays were drawn from those used in previous marine chart and radar studies (Donderi, 1999 and Experiment 1 of this report). All 96 of the displays were unique and there was no duplication of chart or radar displays in either the same or the different condition. Same and different chart displays were selected to represent about equally the four harbour areas. Each mismatched radar display was from the same harbour area as its chart display, but from a non-overlapping part of the harbour.

Ten observers rated the chart and the radar display in each pair for complexity, and then rated the two displays for dissimilarity. Ratings were made on a 1 – 9 scale, with larger numbers indicating greater complexity and greater dissimilarity. Each pair was rated separately, and the pairs were presented in a different random order to each observer. [2]

Procedure

Observers were assigned in alternation to either the ad lib or the fast condition. The S and D key locations were reversed for successive observers in each condition. Observers read on-screen instructions explaining the experimental task, and the experimenter answered any questions regarding the task. A warm-up set of four chart-radar pairs was presented under the appropriate experimental conditions, and the data were discarded. Then the set of 48 experimental pairs was presented in an independently randomized order for each observer. The ad lib condition took about 30 minutes to complete and the fast condition took about 20 minutes to complete.

Results

Ad Lib condition

Figure 13 shows that the mean of log chart VT increased with an increase in LJPG, accounting for a significant 18 percent of the VT variance, and that the VT maximum increased more rapidly as a function of LJPG than did the average viewing time.

There were four possible independent influences on the proportion of correct responses: chart viewing time, chart LJPG, radar viewing time and LJPG. Table 4 shows the correlations among these variables across all of the display pairs.

---

[2] The scaling study was carried out by Graeme Thomas and Tanya Peixe as part of an honours research project in the McGill University department of psychology. The single – pair presentation method was used because it allowed us to obtain dissimilarity ratings between each chart – radar pair.

In order to evaluate the independent effects of the VT and LJPG variables, partial correlation coefficients were calculated between the proportion correct and two of the predictor variables: the mean log of radar VT with radar LJPG partialled out, and radar LJPG with mean log radar VT partialled out. The partial correlations are shown in Table 5.

Table 4. Correlations among Variables across Display Pairs for the Ad Lib Experiment

	Chart VT	Radar VT	Chart LJPG	Radar LJPG	Proportion correct
Chart VT	1.00	***.29***	***.43***	-.02	.07
Radar VT	***.29***	1.00	.12	-.13	***-.40***
Chart LJPG	***.43***	.12	1.00	-.02	.05
Radar LJPG	-.02	-.13	-.02	1.00	***.30***
Proportion correct	.07	***-.40***	.05	***.30***	1.00

* Correlations significant at $p < .05$ indicated in boldface.

Table 5. Partial correlation coefficients predicting mean accuracy from JPEG and latency variables.

Proportion correct and:	Raw correlation	Partialled out	Partial correlation	Significance
Radar VT	-.40	Radar LJPG	***-.47***	$p < .001$
Radar LJPG	.30	Radar VT	.28	$p < .07$

* Correlations significant at $p < .05$ indicated in boldface.

Proportion correct was correlated with both radar VT and radar LJPG, but the effect was stronger for radar VT and partial correlation coefficients showed that only the effect of radar VT independent of radar LJPG was significant (Table 5). Longer viewing times were associated with fewer correct responses, while larger radar LJPG produced marginally significant higher proportions of correct responses (Figure 14). The results did not differ significantly for same or for different pairs.

The dissimilarity of the chart and radar pairs was judged using magnitude estimation scaling (Stevens 1975). Figure 15 shows the relationship between the geometric mean of the judged dissimilarity of the chart-radar pairs, the mean log radar VT and the proportion of correct responses. As dissimilarity increases, it takes longer to make a decision about the same pairs but less time to make a decision about the different pairs, and as dissimilarity increases, decisions to same pairs become less accurate while decisions to different pairs become more accurate (Figure 18).

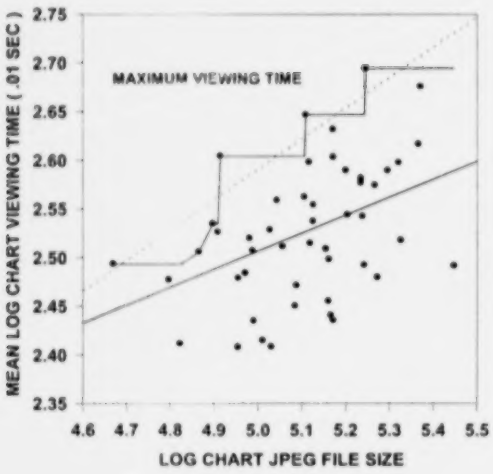


Figure 13. **Experiment 3: Log chart viewing time as a function of log chart JPEG file size.**

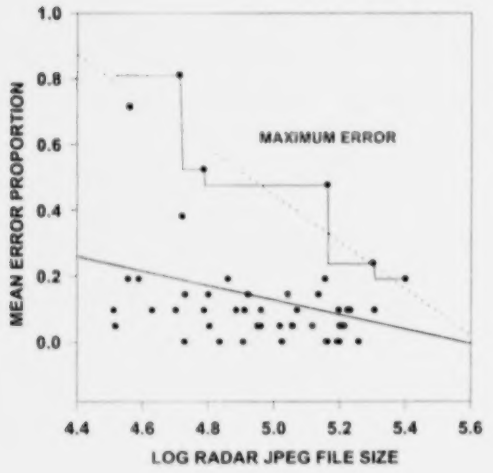


Figure 14. **Experiment 3: Mean error proportion as a function of log radar JPEG file size.**

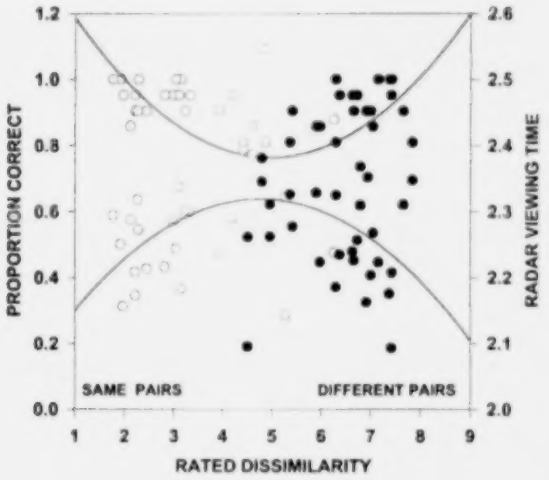


Figure 15. **Experiment 3, ad lib condition: Effect of dissimilarity on proportion correct and radar viewing time for same and different pairs of experiment 3.**

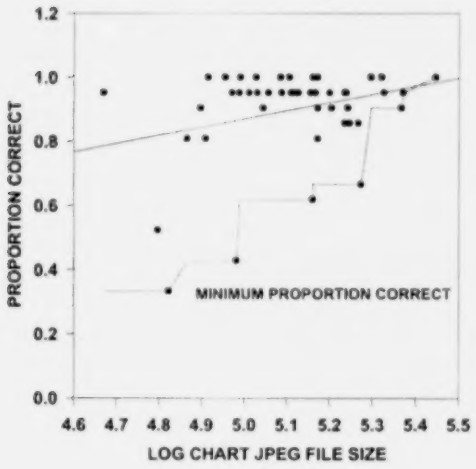


Figure 16. **Experiment 3, fast condition: proportion correct as a function of log chart JPEG file size.**

Fast condition

The performance measure in the fast SD task was proportion of correct responses per display pair, regressed against three predictor variables: the chart LJPG, radar LJPG and an index marking whether the display pair was matched (correct "same" response) or mismatched (correct "different" response). All of the independent variables except the display pair index contributed significantly at $p < .05$ to a prediction equation that accounted for 28 percent of the variance of the proportion of correct responses, as shown in Table 6.

Table 6. Fast Same-Different Experiment: Standardized Regression Coefficients and Partial Correlation Coefficients relating Probability of Correct Response to Each Independent Variable

Variable	*β*	*Partial r*	*p <*
Intercept			***.0037***
Pair type	-0.42	0.24	.1207
Log chart JPEG	12.13	0.31	***.0457***
Log chart JPEG ^{2}	-11.86	0.30	***.0505***
Log radar JPEG	15.60	0.39	***.0110***
Log radar JPEG^2	-15.42	0.39	***.0116***
Mean dissimilarity	-1.92	0.32	***.0394***
Mean dissim^2	2.16	0.37	***.0154***

Inspection of Figures 16 and 17 shows that the significant quadratic term in the regression of proportion correct on chart LJPG was caused by the ceiling effect of a maximum proportion correct at high LJPG values. The significant quadratic term in the proportion correct regression on radar LJPG may be due to a real quadratic effect: proportion correct appears to decrease at higher values of radar LJPG.

As in the ad lib condition, performance varied as a function of dissimilarity depending on whether a display pair was same or different: same pair decisions were more accurate for high values of chart-radar pair dissimilarity, while different pair decisions were more accurate at lower values of chart-radar pair dissimilarity, and performance was worse overall at intermediate levels of dissimilarity, as shown in Figure 18.

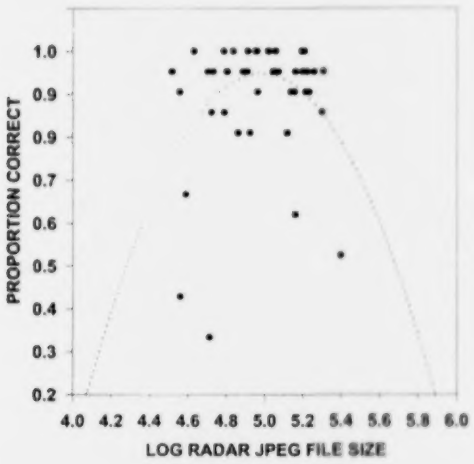


Figure 17. **Experiment 3, fast condition: Proportion correct as a function of radar log JPEG file size.**

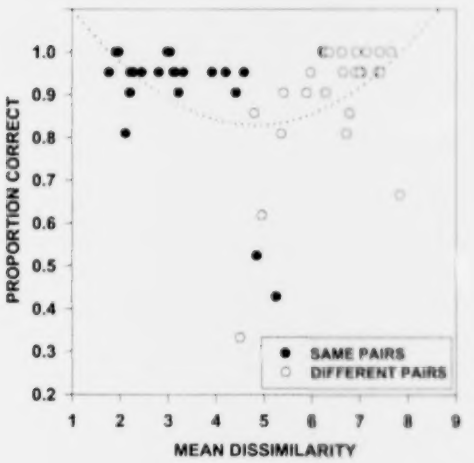


Figure 18. **Experiment 3, fast condition: Proportion correct as a function of mean chart-radar dissimilarity.**

Discussion

When a marine electronic chart display is observed freely in anticipation of a subsequent matching task, the observation time increases with an increase in chart LJPG. This evidence from the ad lib same-different experiment confirms the search experiment findings that LJPG is related to display performance. Under unlimited-time viewing conditions the proportion of correct responses was a decreasing function of radar display VT and an increasing function of the radar LJPG, while chart LJPG had almost no effect.

A simple interpretation of the VT relationships is that chart VT increased as the information load in the chart increased, and since there was no time pressure, the available VT was sufficient to allow the observer to perceive and remember the relevant chart information up to the limit of his or her ability. More information in the radar image associated with a larger radar LJPG made it more likely that a match or mismatch could be resolved on the basis of the adequately stored information from the chart. With less radar information, it would be harder to recognize a match or a mismatch with the chart information. Increased radar VT resulted from the uncertainty caused by a failure to remember the information in the chart and was thus associated with a lower proportion correct, while a higher proportion correct associated with larger radar LJPG resulted from the availability of more radar information to facilitate a match with the chart information.

In the fast same-different experiment, the proportion of correct responses increased as the chart LJPG increased, but was an inverted U-function of radar LJPG. Again the task was to match the chart and radar information. A larger chart LJPG meant more chart information to match against the radar information, and as much as was available was perceived and remembered during the short exposure duration. A larger radar LJPG meant more radar information to perceive and remember, and both the chart information and the radar information had to be perceived and remembered. It appears that the total capacity of the observer to store and compare information from two successive short exposures was exceeded as radar LJPG increased beyond an upper limit.

EXPERIMENT 4: THE EFFECT OF STRUCTURAL AND SEMANTIC INFORMATION ON VISUAL SEARCH PERFORMANCE IN CHART SYMBOL DISPLAYS

Abstract

Observers searched 48 displays containing 36 symbols to find one, two or three target symbols. The target symbols were either triangles or parallelograms; the background symbols were other simple geometric shapes taken from a collection of chart symbols. The displays differed in the number of different symbol types displayed and in the number and arrangement of groups of symbols on the display. Decision latency was an increasing function of the number of different symbols displayed, the number of symbol groups on the display, and the constraints placed on symbol group positions across the display. The results show that a structural analysis of alphanumeric display complexity can also be applied to non-alphanumeric symbolic displays.

Introduction

Tullis (1983) reviewed the factors that influenced performance efficiency with alphanumeric visual displays. He reported that the optimum global display density for a standard 80 column x 24 row character display screen was between 30 and 40 percent filled. He also reviewed the concept of display grouping and reported on the derivation of grouping measures based on the distinction between global and local display density. If the distance between adjacent horizontal characters is called 1, the distance between adjacent vertical characters is called 2, and the average distance between all adjacent characters measured this way is N then all characters separated by less than 2N spaces form a local group. Using this formula, the groups in a display can be unambiguously defined by comparing local density (distance between characters) with global density (average distance between characters). An example of a grouping calculation from Tullis' paper is shown in Figure 19.

Tullis then reviewed and refined several information-theoretical measures of display organization that were based on display groupings, and so ultimately on the distinction between global and local display density. The simplest of these measures was calculated by identifying the coordinates of the upper-left-hand corner of each group on the display using the 80 columns and 24 rows as horizontal and vertical coordinates. Then the number of different horizontal and vertical group coordinates was summed over the entire set of groups, and an information measure was derived from the probability (summed over all of the groups) that a particular horizontal or vertical coordinate defined the upper-left-hand corner of any group. The probability for a particular coordinate value is low, and the resulting information value is high, when the group coordinates are distributed randomly across the display.

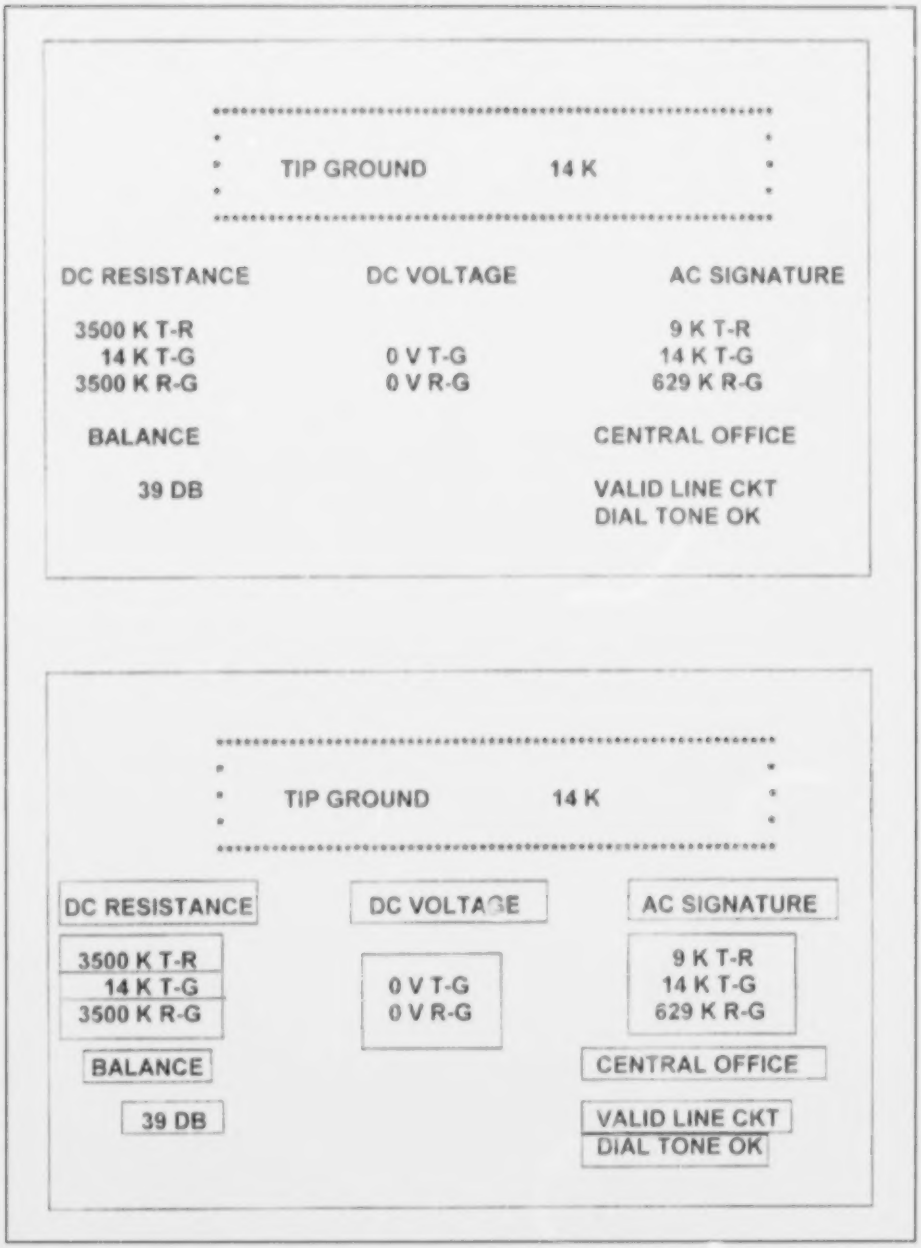


Figure 19. **Experiment 4: Alphanumeric screen (above) with symbol group rectangles outlined (below). Adapted from Tullis (1983).**

This produces a high measure of complexity, which is the same thing as a high structural information content for the display. The opposite situation, when the different groups are arranged in rows and columns, reduces the information content of each coordinate, and, summed over all of the groups, reduces the measured complexity (structural information content) of the display.

Tullis' structural display complexity measure is based entirely on the spacing of characters on the display and it is unrelated to the actual characters that comprise the display. Because it does not depend on the meaning of the display characters, it could be called a syntactic complexity measure, by analogy to the distinction between grammar (syntax) and meaning (semantics) in language. Following this analogy, the grammar of a display is interpreted as its grouping structure, while the meaning of a display is interpreted as the symbolic content of the display groups. The display on the right side of Figure 20 has the same Tullis structural complexity as the display on the left side but the two groups have an entirely different meaning and therefore an entirely different semantic complexity. Tullis recognized that the structural display complexity measure was incomplete because it ignores the effect of meaning (the semantics) on display complexity when he wrote "The elimination [from consideration] of these semantic characteristics is not meant to deny their importance; they clearly are critical to the usability of a display. However, semantic characteristics represent a different realm from the spatial characteristics considered in this paper" (Tullis, 1983, p 661).

DRDC – TORONTO RESEARCH PROJECT NOVEMBER 2002	XQJP – MLZDRPW NBJTLARD PMOJERT XMEUADFI 8170

Figure 20. **Two alphanumeric displays with identical structural content and different semantic content.**

Method

The experiment was designed to study the simultaneous effect on display performance of differences in both semantic and structural complexity. The task was target search, and the logic was that it will be harder to find a target in complex displays than in simple displays. Following Tullis, the structural (syntactic) component of display complexity was measured by character grouping, and the meaning (semantic) component was based

on the variety and distribution of characters within the groups. The non-letter characters that formed the displays were drawn from a set of chart symbols.

The experiment required the observer to search each display in order to decide which of two targets was present (a target was always present), and also to decide how many (1, 2 or 3) of that target were present. Multiple targets were added to make it less likely that an observer could complete a serial self-terminating search before attending to all of the information in the display. After finding the maximum of three targets in a display the observer could terminate the search, but only about one-third of the displays had three targets; these were randomly distributed across the displays, and for two-thirds of the displays, the display would have to be exhaustively searched before reporting that only one or two targets were present. The alternative of putting more targets in each display was rejected because it might be hard to search for more targets accurately under time pressure. A larger number of targets might lead observers to compromise accuracy (searching the entire display) for speed (searching some of the display carefully and some hastily if at all). The goal was to measure the effect of the complete assimilation of display information on the decision time to an accurate, not a hasty, decision.

Stimuli

Each display contained thirty-six chart symbols drawn from a set of ten whose discriminability had been previously established (McFadden *et al*, 1997). Eight of the symbols were used to create the background and two were used as targets. One, two or three of the items in each display were the same target and the rest were background. Both the target, the number of times it was presented and the target display position or positions were varied randomly across the displays.

There were forty-eight different displays. Sixteen displays each consisted of a randomly ordered selection from eight background symbols plus the targets, another sixteen displays each consisted of a randomly ordered selection from four of the background symbols plus the targets and another sixteen displays consisted of multiple examples of only one randomly selected background symbol plus the targets. Thus for each display the target symbol was randomly selected from two alternatives, the number of targets was randomly selected as 1, 2 or 3, and background symbols were selected and randomly ordered from either 1, 4 or eight of the symbol types.

All of the symbols in each display were arranged in groups. Each group contained either six or twelve symbols. The groups of twelve had four symbols in three rows. The groups of six had three symbols in two rows. The twelve-symbol groups were 4 cm wide by 3 cm high on the display and subtended a visual angle of 7.6 deg horizontally and 5.7 deg vertically. The six-symbol groups were 3 cm wide by 2 cm high on the display and subtended a visual angle of 5.7 deg horizontally and 3.8 deg vertically. Within each set of sixteen displays with 1, 4 or 8 background symbols, four displays contained three groups of twelve for a total of three groups, four contained two groups of twelve and two groups of six for a total of four groups, four contained one group of twelve and four groups of

six for a total of five groups, and four contained six groups of six items for a total of six groups.

In order to arrange the groups, the display surface (22 cm horizontally by 22 cm vertically) was divided into thirty-six equal areas (six areas across and six areas down). The areas were numbered with horizontal (row) and vertical (column) coordinates one to six. Using this area grid, the horizontal and vertical coordinates of the upper left corner of each group were assigned either randomly or randomly with constraints. Under random assignment, two sets of numbers from one to six were drawn without replacement to determine the upper left hand horizontal and vertical coordinates for the area occupied by each group. Separate pairs of coordinates were drawn for each group. Under constrained random assignment, fewer horizontal and vertical coordinates were drawn than there were groups, and groups were placed within the constrained set of coordinates. For example, for displays with five groups, the unconstrained random assignment drew five different coordinate pairs for the five groups. The constrained random assignment drew three horizontal coordinates and two vertical coordinates, and the five groups occupied all but one of the six resulting coordinate pairs. The unfilled coordinate pair was chosen at random. A similar process was followed to locate the unconstrained and constrained groups for the displays with 3, 4 and six groups. Two displays are shown in Figure 21.

Table 7. Information analysis of the Displays used in Experiment 4.

N (groups)			3		4		5		6	
Type (groups)			3 x 12		2 x 12, 2 x 6		1 x 12 , 4 x 6		6 x 6	
Position (groups)			R	C	R	C	R	C	R	C
N (coordinates)			3 x 3	2 x 2	4 x 4	3 x 2	5 x 5	3 x 2	6 x 6	4 x 2
H(ST)			3.16	2.00	4.00	2.58	4.64	2.58	5.16	3.00
NS	NT	H(SE)	H(Total)							
1	1	6.62	9.78	8.62	10.62	9.20	11.26	9.20	11.78	9.62
	2	11.16	14.32	13.16	15.16	13.74	15.80	13.74	16.32	14.16
	3	14.90	18.06	16.90	18.90	17.48	19.54	17.48	20.06	17.90
4	1	76.54	79.70	78.54	80.54	79.12	81.18	79.12	81.70	79.54
	2	79.05	82.21	81.05	83.05	81.63	83.69	81.63	84.21	82.05
	3	80.82	83.98	82.82	84.82	83.40	85.46	83.40	85.98	83.82
8	1	117.32	120.48	119.324	121.324	119.90	121.96	119.90	122.48	120.32
	2	118.58	121.74	120.584	122.584	121.16	123.22	121.16	123.75	121.58
	3	119.09	122.25	121.088	123.088	121.67	123.73	121.67	124.25	122.09

H(ST) = structural information, NS = number of symbols, NT = number of targets, H(SE)= semantic information, H(Total) = total information, R = random groups, C = constrained groups.

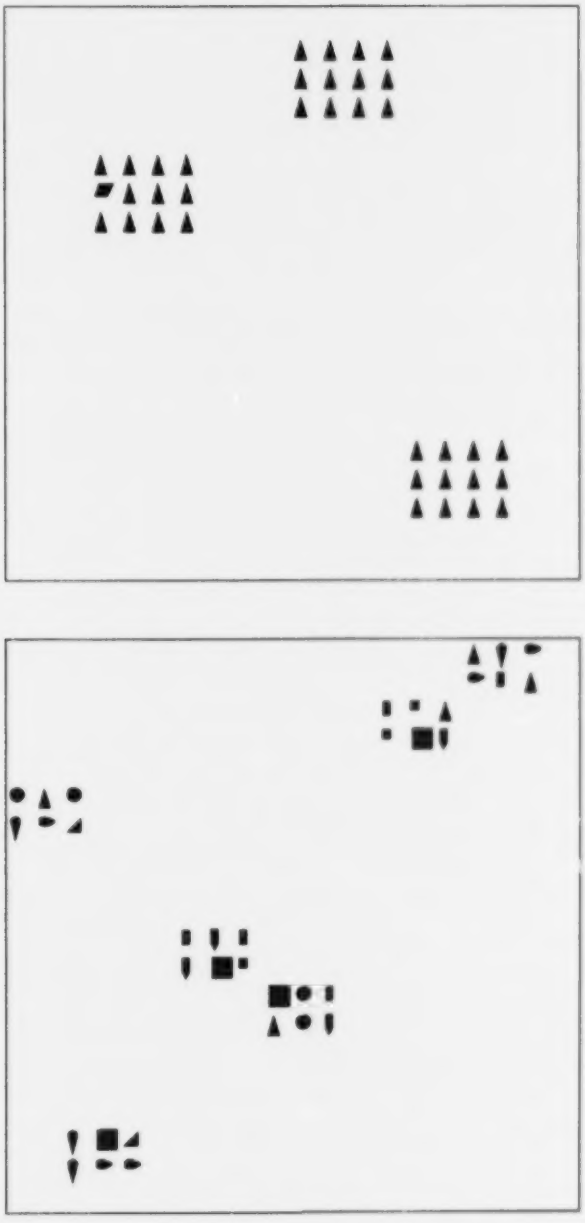


Figure 21. **Experiment 4: Structurally and semantically simple (top) and complex (bottom) chart symbol displays.**

Display information measures

Tullis' (1983) structural information measure depends on the number of groups in each display and on the constraints imposed on the group positions. The semantic information content of each display was based on the number and distribution of symbols across the display, independently of the group structure of the display. The information measurements used are explained in Quastler (1954) and Attneave (1959). The structural information, the semantic information, and the information sum for each display type, type, are shown in Table 7.

Instructions

Observers were told to decide which of two targets (triangle or parallelogram) was in each display, and how many (1, 2 or 3) there were. They were asked to respond as quickly as possible consistent with accuracy. They pressed one of two computer keyboard keys (the bracket keys, each re-labelled with an image of one of the targets) to show which target they saw in the display. As soon one of those keys was pressed, the display disappeared. Decision latency was measured in hundredths of a second from display onset to display disappearance. Then the display screen presented a number box for entering how many targets were seen. The box accepted entries of 1, 2 or 3 targets. After a number had been entered, the display cleared and the observer was instructed to press the "enter" key when ready for the next display. Two seconds later the next display appeared.

Presentation conditions

Observers sat about 30 cm from a 17" flat CRT display. Experimental instructions were shown on the display screen before the displays were presented. Each observer started with a practice task consisting of three displays that were not part of the experimental set, and data from these displays was discarded. Then each of the 48 experimental displays was presented in an independently randomized order. The positions of the keys labelled with the triangle and parallelogram were reversed between observers. The entire experiment, including instructions, explanations and warm-up, took about twenty minutes.

Results

The mean decision time averaged over 44 observers, the standard deviation of decision time and the proportion correct for each of the 24 combinations of symbol (1, 4 or 8), group (3, 4, 5 or 6) and constraint (random or constrained) are given in Table 8.

Table 8. Mean and SD of Decision Latency, and Proportion of Correct Responses, for Number of Background Symbols, Number of Groups, and Random or Constrained Group Coordinates in the Symbol Search Experiment

Groups	Coordinates	Number of background items								
		1			4			8		
		M	SD	Correct	M	SD	Correct	M	SD	Correct
3	R	217	92	98	420	224	92	526	238	90
	C	247	104	99	403	159	72	579	419	86
4	R	217	88	98	482	265	97	466	218	88
	C	193	76	100	472	232	68	594	272	90
5	R	246	99	48	443	194	85	562	251	60
	C	244	90	100	416	161	97	509	222	90
6	R	249	82	100	510	224	99	559	273	83
	C	204	87	98	456	212	48	552	254	82

The standard deviations of decision times were plotted against the mean times for each of the 24 combinations of symbol, group and constraint, and for right and wrong decisions. There was significant linear correlation ($r = 0.92$) between the mean and the SD. Decision times were log transformed (LDT) for subsequent presentation and analysis.

An analysis of covariance was used to assess the significance of differences in LDT among symbols (1, 4 or 8), groups (3, 4, 5 or 6), coordinates (random or constrained), and their interactions. Observers were the random factor and there were four covariates: the number of targets presented per display (1, 2 or 3), the order of display presentation (1 to 48), and the linear and the quadratic LJPG terms for each display. In all analyses reported here the main effect sums of squares were adjusted for interactions, so that significant main effects remain significant despite significant higher-order interactions involving the main effect factors.

The main effects of symbols ($F [2,2020] = 393.6$, $p < .0001$), groups ($F[3,2020] = 3.82$, $p < .01$) , and constraints ($F[1,2020] = 4.59$, $p < .05$) were all significant at $p < .05$, as were the two-factor interactions of symbols x groups ($F[6,2020] = 5.30$, $p < .0001$), groups x constraints ($F[3,2020] = 22.2$, $p < .0001$), groups x decisions ($F[3,2020] = 9.98$, $p < .0001$) and the three-factor interactions of symbols x groups x constraints ($F[6,2020] = 4.46$, $p < .0005$), symbols x groups x decisions ($F[6,2020] = 4.08$, $p < .0005$), and groups x constraints x decisions ($F[3,2020] = 5.35$, $p < .005$). The covariates of number of targets ($F[1,2020] = 27.03$, $p < .0001$) and presentation order ($F[1,2020] = 231.08$, $p < .0001$) also significantly influenced LDT.

Figure 22 illustrates the main effects of symbols and groups on LDT. Number of symbols has the largest effect, with increasing numbers of symbols leading to longer LDT. More groups also produce longer LDT. The interaction between symbols x groups is shown by the deviation of the individual data points from the regression line relating number of groups to decision latency at each level of symbols.

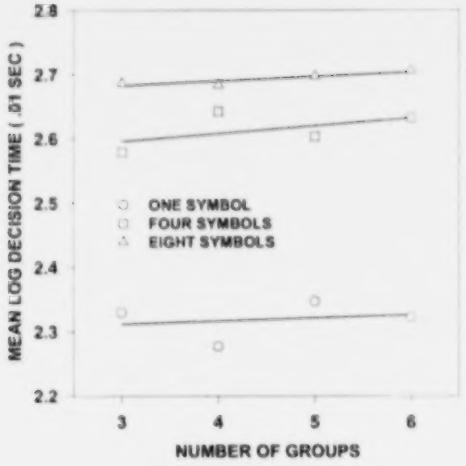


Figure 22. **Experiment 4: Mean log decision time as a function of the number of groups and the number of symbols.**

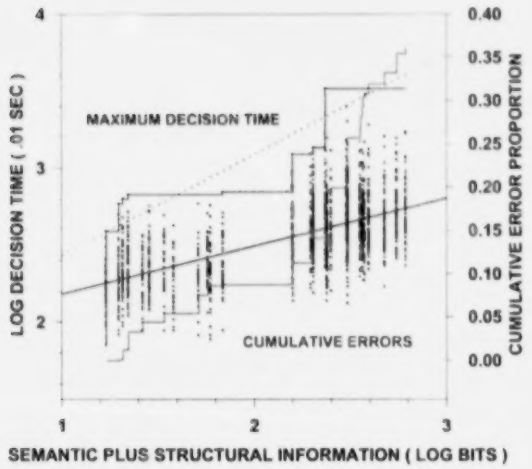


Figure 23. **Experiment 4: Mean log decision time as a function of the sum of semantic and structural information for all responses.**

The three-way interactions of symbol x group x constraints and symbol x group x decisions mean that the two-way symbols x groups interaction pictured in Figure 22 is itself perturbed, independently, by each of the two other factors. In other words, the symbols x groups interaction is different depending on whether the decision was right or wrong and whether the display was random or constrained. But none of the interactions compromise the conclusion that each one of the main effects of symbols, groups and constraints have a significant independent influence on LDT.

A separate analysis of covariance was carried out on the proportion of correct decisions for each combination of symbol, group and constraint. The main effects sums of squares were adjusted for interactions, so that significant main effects remain significant in the presence of significant interactions. The main effects of symbols ($F[2,3068] = 39.46$, $p < .0001$) and groups ($F[3,3068] = 9.06$, $p < .0001$) were significant at $p < .05$, and all of the interactions, including those with the non-significant main effect of constraints, were significant at $p < .0001$ with a minimum F over all of the interactions of 21.68. The covariate of presentation order had a significant effect ($F[1,3068] = 3.97$, $p < .05$).

Figure 24 plots LDT against LJPG for the 48 displays. The clearly non-linear relationship was confirmed by a regression analysis of LDT against the linear and quadratic LJPG terms, which showed significant and equal contributions as indicated by standardized regression coefficients of 92.13 for the linear term -92.12 for the quadratic term at $p < .001$. The quadratic fit was the same for both right and wrong decisions. No relationship, either linear or quadratic, was found between JPEG file size and the proportion of correct responses.

Information measures

The sum of the structural and semantic display information was plotted against LDT for each response to each display in Figure 23, which shows a linear increase in LDT with an increase in the information sum over the entire range, a faster increase in the maximum than the mean LDT value with increased information sum, and a delayed rise in the cumulative proportion of errors until an information sum threshold is passed. For the symbolic displays in this experiment, the sum of semantic and structural information acts like LJPG in the search experiments with marine electronic displays and colour photographs.

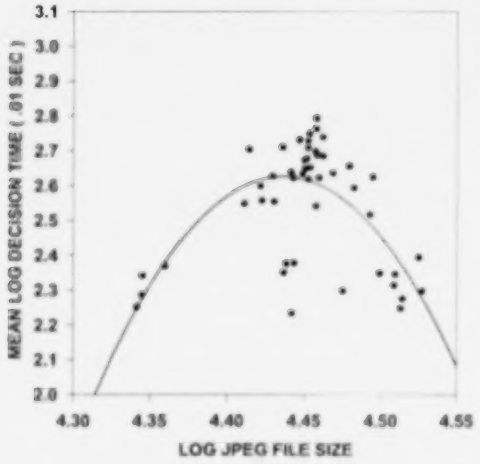


Figure 24. **Experiment 4: Mean log decision time as a function of log JPEG file size.**

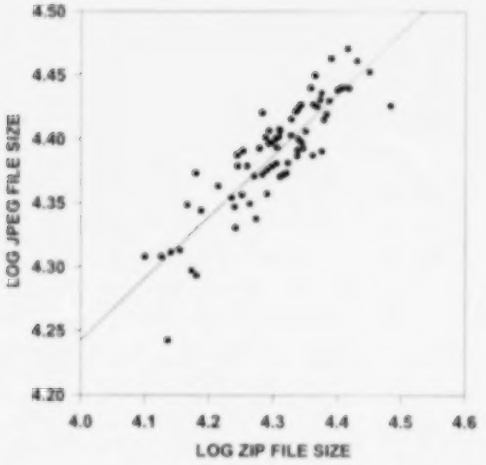


Figure 25. **Experiment 5: Log JPEG file size as a function of log ZIP file size for the chart symbol figures.**

Discussion

According to Tullis, the structural information in an alphanumeric display is based on the number and positional variety of symbol groups in the display. With chart symbols instead of letters or numbers forming the display, log decision latency increased with an increase in the number of symbol groups. Log decision latency was greater for groups with random coordinates than for groups with constrained coordinates. Increasing the number of groups and randomizing the coordinates increase both the structural information content of the display and the decision latency to the display. LDT was an increasing linear function of the sum of the semantic and structural information in each display.

The linear and quadratic components of LJPG were used as covariates in all of the analyses described above. Therefore the significant effects of groups, symbols and constraint rules are independent of LJPG. However LJPG had a strong effect on LDT as shown in Fig 24. The non-linearity is due to the very large JPEG file sizes for some single-symbol displays that consist of chart symbols with curved outlines, as shown in Figure 21. The entire cluster of points in the lower right-hand corner of Fig 24 consists of single-symbol displays composed of the items like the curved chart symbols shown in Figure 21. Although these single-symbol displays are quickly assimilated by the human eye and brain, the JPEG algorithm generates a surprisingly longer file for this subset of the single-item displays. Other rectangular or triangular single-symbol displays like those clustered around the lower left-hand corner of Fig 24 are both easy for the human to process and easy for JPEG to process: they produce fast decision latencies and small JPEG files. This suggests that for displays composed of separate symbols, JPEG file size might not relate to subjective complexity and its concomitant in objective performance for displays composed of discrete symbols as well as it does for the marine chart, radar and overlay displays. However, Tullis structural information and semantic information calculated from the distribution of symbols over the display provide an alternative measure of complexity for this type of display.

EXPERIMENT 5: VISUAL SEARCH OF SYMBOLIC DISPLAYS II: MORE DISPLAYS AND FEWER SYMBOLS

Abstract

Eleven observers searched a set of 72 displays composed of 24 chart symbols to find either a triangle or parallelogram target among the remaining different background symbols. There were either 1, 2 or 3 different background symbols arranged in six groups of four, twelve groups of two or individually according to one of two rules: either the x and y coordinates for each symbol (or group) were independent, or the groups were constrained to fall in rows and columns. Changing the number of symbols varied the semantic information while changing the number of groups and the placement constraint varied the structural information. Log decision time (LDT) increased with an increase in rated complexity and semantic information content. Another group of twelve observers rated each display for complexity, which increased with an increase in semantic and structural information.

Introduction

In experiment 4, displays constructed from chart symbols were used to show that semantic information content (related to the number of symbols that appeared in the display) and structural information, related to the number and spatial distribution of groups of symbols in the display, predicted log decision time (LDT) to find and count the target symbols in the display. There was also an inverted U-relationship between LJPG and LDT. The inverted-U function relating LJPG to the information and performance measures was unexpected. Experiment 5 was carried out to learn whether the results of experiment 4 could be replicated. There were some differences in procedure. While experiment 4 required observers to both identify and count the targets (from 1 to 3) in each display, experiment 5 only required them to identify the single target per display. Otherwise, the task and the method of display presentation were the same across the two experiments. An independent group of observers rated the judged complexity of each of the new displays.

Method

Displays

Each display contained 24 symbols drawn from the set of chart symbols studied by McFadden *et al* (1997). One symbol in each display was the target: either a triangle or a parallelogram chosen at random. Eighteen of the displays had either one, two, three or four different symbols selected at random from the rest of the chart symbol set to form the background. If there was more than one background symbol, they were divided as equally as possible, given the odd number of background items, between each symbol. In 24 displays the symbols were arranged in six groups of four in a square, in 24 displays there were twelve groups of two in a horizontal line, and in 24 displays there were 24 "groups" of one symbol each. The horizontal and vertical coordinates of the top left corner of each group were chosen by one of two rules: random or constrained. The

random rule meant that each coordinate was chosen at random, independently of every other (except to avoid overlapping groups). Thus the groups could appear anywhere on the display. The constrained rule meant that the total number of horizontal times vertical coordinates equalled the total number of groups, and the number of horizontal and vertical coordinates was as close to equal as possible. For example, for six groups there could be either three horizontal and two vertical coordinates or two horizontal and three vertical coordinates. Similarly for twelve groups there would either be 4 x 3 or 3 x 4 (horizontal and vertical) coordinates, and for 24 groups, either 6 x 4 or 4 x 6 horizontal and vertical coordinates. This meant that under the constrained rule, the groups were always arranged in a rectangular array with random spacing between the rows and columns of the array. Half of all of the displays had groups distributed by the random rule and half had them distributed by the constrained rule. Two displays from the set used in Experiment 5 are shown in Figure 26.

As in experiment 4, the purpose of these rules was to systematically vary the two display information components, semantic and structural information, over all of the displays. Semantic information was varied by changing the number of background symbols, taking into consideration the information contributed by the single target symbol. Structural information was varied by changing the number of symbol groups and the rules about how to place those groups. Table 9 shows the distribution of semantic and structural information across the display groups in the experiment.

Table 9. Semantic and Structural Information Measures for the displays in Experiment 5.

Groups		6		12		24	
	Rule	R	C	R	C	R	C
Semantic information (bits)		*Structural information(bits)*					
Symbols							
1	4.58	5.16	2.58	7.16	3.58	9.16	4.58
2	6.64	5.16	2.58	7.16	3.58	9.16	4.58
3	9.45	5.16	2.58	7.16	3.58	9.16	4.58
4	12.78	5.16	2.58	7.16	3.58	9.16	4.58

Presentation

Instructions were the same as in experiment 4 except that observers did not enter the number of targets seen in each display (it was always 1). Instead, they entered a confidence rating for each decision. The presentation conditions were the same as in experiment 4.

Complexity Rating

A different group of twelve observers rated each of the displays for complexity. The displays were presented in random order on a computer screen for as long as required by

the observer. The observer was instructed to look at each display and to rate its complexity using any range of numbers greater than zero, with larger numbers indicating greater complexity. If one display appeared twice as complex as another, the observer was instructed to give it a complexity rating twice as large. A set of six practice displays were presented first, and the practice data was discarded.

Results

JPEG and ZIP file sizes were plotted against each other in Figure 25, which shows a strong linear relationship in log-log coordinates and that ZIP and JPEG compressed file lengths are about equal. This confirms the fact that over a large range of display types, ZIP and JPEG file compression algorithms give monotonically related results, but that JPEG compresses better only with more complex images.

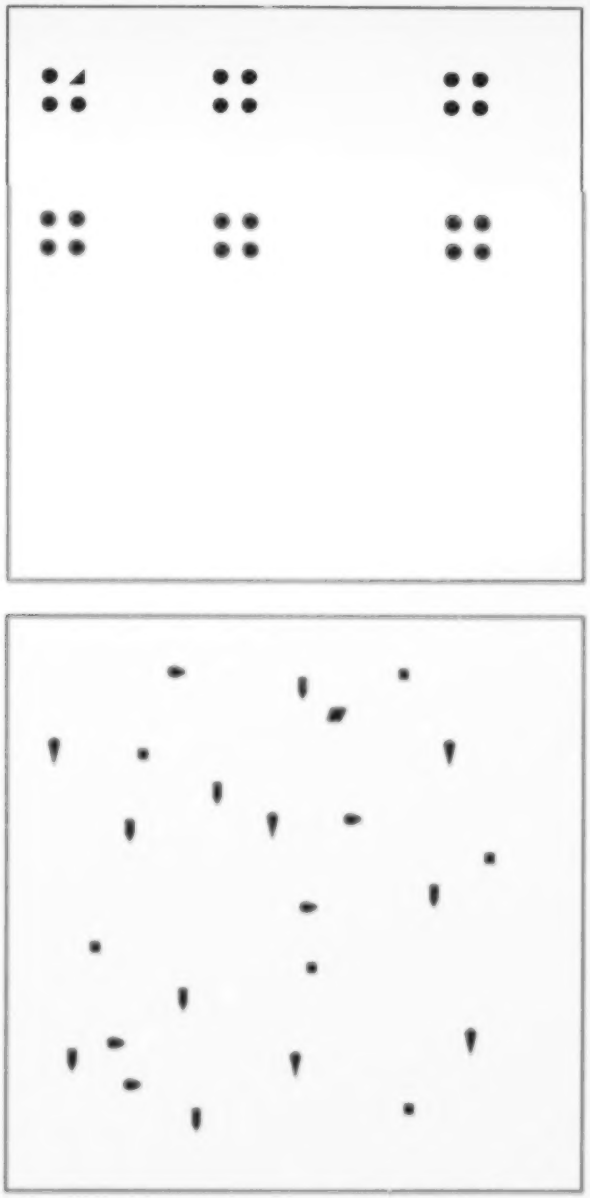


Figure 26. Experiment 5: Simple (top) and complex (bottom)) chart symbol display

Table 10 gives the LDTs and their SDs for the display design factors of groups, constraint rule and background symbols.

Table 10. Log decision time and (SD) for Display Design Factors: Symbols, Groups and Rules.

Groups	*6*		*12*		*24*	
Rules	*C*	*R*	*C*	*R*	*C*	*R*
Symbols						
1	2.04 (.17)	2.12 (.27)	1.98 (.16)	2.05 (.16)	2.07 (.20)	2.09 (.25)
2	2.15 (.24)	2.16 (.21)	2.12 (.17)	2.17 (.23)	2.06 (.17)	2.10 (.26)
3	2.17 (.18)	2.14 (.22)	2.17 (.20)	2.18 (.32)	2.18 (.19)	2.08 (.22)
4	2.21 (.23)	2.23 (.34)	2.22 (.27)	2.14 (.20)	2.22 (.18)	2.20 (.25)

Only 0.007 of all trials produced wrong responses. Analysis of variance of the display design factors and order of presentation showed no significant influences of any of the display design factors on errors. Analysis of variance of the LDTs showed a significant effect of number of symbols [$F(3,710) = 26.75$, $p < 0.0001$], a significant interaction between number of groups and number of symbols [$F(6,710) = 2.23$, $p < 0.04$], a trend towards significance for the interaction between constraint rule and number of symbols [$F(3,710) = 2.40$, $p < 0.07$], significant differences among observers [$F(11,710) = 49.17$, $p < 0.0001$] and a significant effect of order of presentation as a covariate [$F(71,710) = 1.55$, $p < 0.005$].

The semantic and structural information measures described in Table 9 were entered into a regression analysis with LDT and proportion of errors as the dependent variables. Both dependent variables increased linearly with an increase in semantic information (number of background symbols) as shown in Figure 27, but structural information had no significant effect on either LDT or proportion of errors.

Analysis of variance of the effect of display design factors (groups, symbols and constraint rule) on average rated complexity was carried out with order as a covariate. Each design factor had a significant effect at $p < 0.005$. Order had no significant effect and there were no significant interactions among the three design factors.

The average rated complexity of each display was plotted as a function of semantic information and structural information in Figure 28. Regression analysis showed that each information source contributed independently to predict sixty-two percent of average complexity. The best fit equation was: average complexity = 1.09*semantic information + 0.72* structural information + 2.68. The standardized regression parameters were $\beta = 0.71$ for semantic information and $\beta = 0.33$ for structural information. Figure 29 illustrates this relationship.

Average rated complexity, which was measured independently of LDT, predicts LDT. Figure 30 shows the relationship between LDT, mean log complexity, and the cumulative proportion of errors. LDT increases regularly with increases in complexity, but until a complexity threshold 1 log unit is passed, there are no errors.

LDT was regressed against linear and quadratic LJPG terms, and the result showed a trend towards significance ($p < 0.07$). Figure 31 shows that the maximum LDTs occurred in the middle of the LDT range, as in Experiment 4, with low LDTs occurring throughout the range. Figure 32 shows that mean log complexity, which was measured independently of LDT, also has a peak at intermediate values of LJPG.

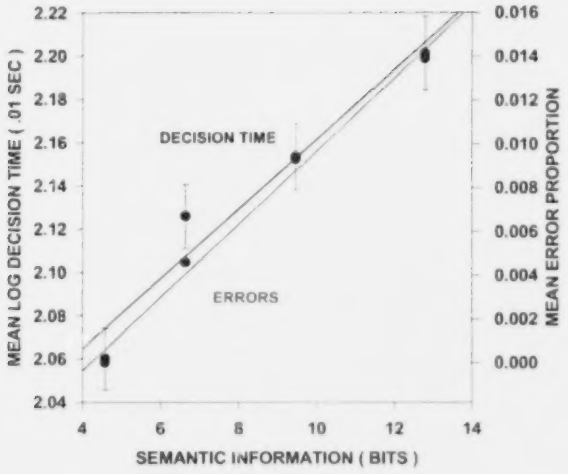


Figure 27. **Experiment 5: Mean log decision time and mean proportion of errors as a function of semantic information in bits.**

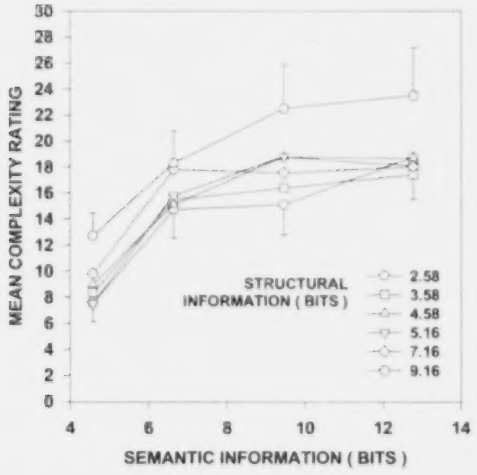


Figure 28. **Experiment 5: Mean complexity as a function of semantic and structural information.**

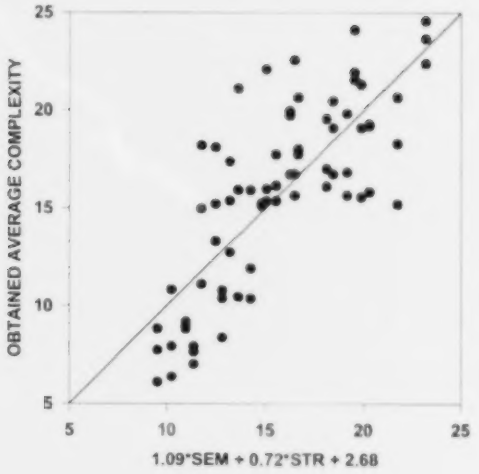


Figure 29. **Experiment 5: Prediction of complexity rating for all displays from semantic and structural display information.**

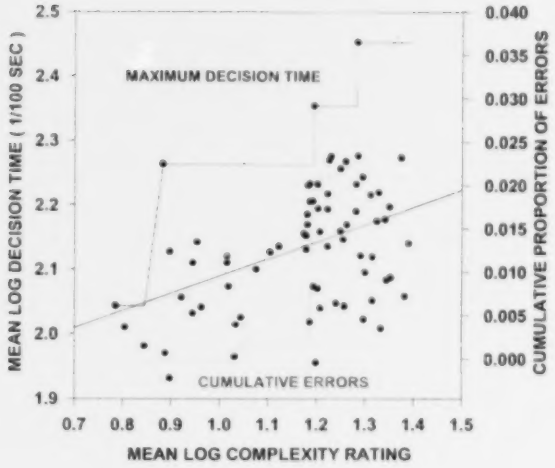


Figure 30. **Experiment 5: Mean log decision time as a function of mean log complexity rating.**

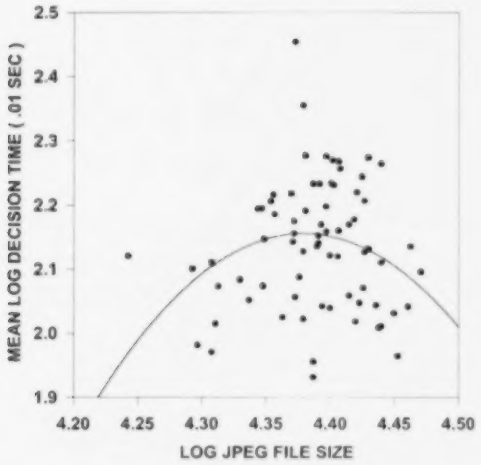


Figure 31. **Experiment 5: Mean log decision latencyas a function of mean log JPEG file size.**

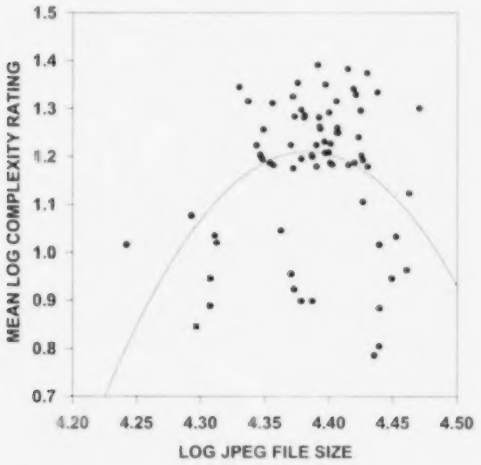


Figure 32. **Experiment 5: Mean complexity rating as a function of mean log JPEG file size.**

SUMMARY AND CONCLUSIONS

Experimental Findings

This section summarizes the results of the five experiments that were completed during the course of this project.

Experiment 1: Target search in chart, overlay and radar displays.

Observers saw 53 displays, one-third of each type, and responded by pressing a key to signal which of two targets they saw in the display. The radar displays produced faster responses than the charts which were faster than the overlays, and the overlays were the least accurate. Wrong decisions were more variable and longer than right decisions. The cumulative probability of a wrong response and the cumulative maximum decision latency increased with increasing LJPG. Decision time increased over the entire LJPG range but errors occurred only after LJPG exceeded the threshold of 4.9 LJPG units.

Experiment 2: Visual search of earth image targets.

The displays were 53 aerial colour photos of the earth's surface. The targets were the parallelograms and triangles used in the previous studies, placed at random locations in each image. The displays were presented in random order. The observer signalled the target by a key press which ended the display, and then entered a confidence judgment. A separate group of observers saw the displays in random order and rated the complexity of each display using an open-ended numerical scale. There was an increase in mean LDT as a function of LJPG. The maximum LDT increased with an increase in LJPG, but below that maximum there were both fast and slow responses at every level of LJPG. Mean log complexity increased with increased LJPG.

Experiment 3: Same-Different matching between chart and radar displays.

A chart display was presented for ad lib inspection followed by a radar display also presented for ad lib inspection. The observer had to decide whether the radar displayed the same area as the chart. Half of the radars matched and half did not. The inspection time for charts increased with chart LJPG. The proportion of correct responses was independent of chart inspection time but increased with radar LJPG and decreased with radar viewing time, and these two effects were independent of each other. Another influence on both accuracy and radar inspection time was the subjective dissimilarity of the chart-radar pairs. As dissimilarity increased viewing time increased for same pairs and decreased for different pairs and accuracy decreased for same pairs and increased for different pairs

In the fast condition, the chart and radar displays were presented briefly with a very short inter-display interval. Accuracy increased with increased chart LJPG and accuracy was

greatest at an intermediate radar LJPG. Accuracy was a function of the average dissimilarity between chart and radar image.

Experiments 4 and 5: Semantic and Structural Information in Chart Symbol Displays

In Experiment 4, displays consisting of 36 chart symbols were organized into 3, 4, 5 or six rectangular groups of six or twelve symbols, arranged over the display at random locations or locations constrained by a limited choice of horizontal and vertical coordinates. Either 1, 4 or 8 different symbols were used in each display, in addition to between one and three target symbols (either parallelograms or triangles) that were located at random among the other symbols. Semantic information content was measured using the Shannon measure based on the probability of occurrence of symbols on the display, and structural information was measured independently according to Tullis' adaptation of Shannon information to measure the regularity of the position of blocks of information on the display. Observers saw each of 48 displays in random order. They pressed a key to signal which target was present, and entered a number from 1 to 3 to signal how many were present. Greater structural and semantic information led to longer LDT, and LDT was an inverted U-function of LJPG.

The symbols used in Experiment 5 were the same as those in Experiment 4. There were 72 displays that varied in having 6, 12 or 24 groups of 4, 2 or 1 symbol, respectively. The groups were arranged either at random across the display or were constrained to fall into horizontal rows and vertical columns, and the groups were composed of either 1, 4 or 8 background symbols plus a single target symbol: either a triangle or a parallelogram. Observers saw the displays in random order and inspected them until they made a decision, which they entered by a key press followed by a confidence judgment A separate group of observers saw the displays in random order, and rated the displays for complexity using an open-ended numerical scale. LDT and errors were a function of the semantic complexity of the display. Complexity judgments were a function of both semantic and structural complexity, and each information component produced an additive and independent increment on subjective complexity. Mean LDT was a marginally significant inverted U-function of LJPG.

Conclusions

Search Experiments and JPEG File Size

The experiments show that lossless JPEG file size, an objective measure that is easily obtained for computer-based displays, correlates with performance measures obtained from target search and same-different decision tasks carried out with marine electronic chart, radar and overlay displays and with aerial colour photographs. These results complement the earlier discovery that JPEG file size was correlated with subjective complexity judgments obtained for the same sets of displays.

JPEG file size is a simple, practical and objective display measure that correlates with the subjective experience of display complexity and with objective measures of display performance. The performance measures were realistic: detecting a particular target in a display, and deciding whether a radar image matched a chart display. Both are tasks that a navigator could be called on to carry out with the chart and radar displays in a wheelhouse. Search duration increased, and search accuracy decreased, with increasing JPEG file size. In subjective terms, the more complex the display, the longer it is likely to take to find a target in it and the less accurate the search will be. While this conclusion could have been derived from literally scores of earlier search experiments, the difference is that now it is based on an objective complexity measure – JPEG file size – that is commonplace in the computer world.

The range of the JPEG measures differed across the various types of display. The chart symbol displays of experiments 4 and 5 had JPEG and ZIP file sizes that were an order of magnitude smaller than the chart, radar and overlay displays of experiments 1 and 3, and the colour photo displays of experiment 2 were an order of magnitude larger than the chart, radar and overlay displays of experiments 1 and 3. The bitmap images for the displays covered a narrow range from 1,650 kb to 2,508 kb, so the differences in compressed file size are due more to the coding efficiency for each type of display than to differences in the bitmap size of the basic displays. The chart symbol displays had the smallest LJPGs, and LDT was not monotonically related to LJPG for these displays.

The relationships between LJPG and LDT were similar across the search experiments, but the regression slopes and intercepts varied. The experiments establish only that in several search experiments the relationship between LDT, proportion of correct responses and LJPG has the same general form. The experiments have not shown an invariant relationship across displays and tasks between LJPG and LDT, proportion correct or subjective complexity.

JPEG and ZIP file sizes were highly correlated for all of the display sets where both file types were generated and measured. It would have been repetitive to duplicate all of the JPEG – based analyses carried out here with ZIP – based analyses, but there is nothing necessarily unique about JPEG compression coding. Algorithmic information theory's invariance theorem states that all "universal coding languages" are equivalent, up to a monotonic transformation (Chater, 1996). JPEG and ZIP compression for the display images used in these experiments *are* roughly equivalent up to a monotonic transformation as shown in figures 2, 9 and 25.

Same – Different Experiments

In the fast version of the same – different task, both the chart and the radar JPEG file sizes were related to accuracy, as was the subjective dissimilarity. As chart file size increased, accuracy increased. But for the radar displays, the optimum file size was in the middle of the range. Unlike the electronic chart and aerial photo search results where increased file size always led to longer and less accurate performance, both too little and too much complexity, as measured by JPEG file size, reduced the accuracy of same-

different decisions. This is probably because an upper capacity limit for remembering and comparing the chart and radar information was exceeded when the radar LJPG was relatively large.

Where observers could look at the same – different displays for as long as they wanted, radar viewing time was a measure of their uncertainty and increased as the proportion correct decreased. Performance was better with larger radar LJPG because there was time to assimilate all of the available radar information and compare it to the remembered chart information. A possible reason that chart LJPG did not predict accuracy in the ad lib experiment was that differences in the time spent by observers in viewing the displays allowed individual differences in memory efficiency to compensate for differences in the complexity and information content (as measured by LJPG) of the displays.

Chart Symbol Search Experiments

Two other visual search experiments were carried out in order to clarify some ideas about display complexity developed by Tullis (1983). He reviewed and refined a simple information measure based on the number of different rows and columns used as starting – points for separate display units (closely arranged character strings) in the displays. He found that the simpler the structural complexity of a display, the easier it was to extract information from it. In experiments 4 and 5, the Tullis structural complexity measure was supplemented by a measure of semantic complexity based on the number of different symbols in the display. The display elements in these experiments were chart symbols rather than the alphanumeric characters studied by Tullis. In Experiment 4, the structural complexity measure contributed significantly to predicting LDT, but in Experiment 5, only semantic complexity contributed to that prediction. There was also a non – monotonic effect of LJPG on LDT and complexity unlike the monotonic relationships found in the search experiments. In Experiment 4, there was a subset of relatively simple displays containing only one symbol type that had very high JPEG file sizes, and yet were searched quickly and accurately. The apparent difference between this unusual set of displays and the other apparently equally simple displays was that these displays were constructed from symbols that, to the JPEG algorithm at least, were individually more complex than some of the other symbols. As a result when they formed most of a display, they produced a high LJPG. The inverted U – relationship found in Experiments 4 and 5 may be an artefact of the particular chart symbols used in the displays.

JPEG File Size and Complexity

The theoretical implication of these studies is about the general concept of visual image complexity. While complexity is a concept that might be related to any sensory input, our analysis is here limited to two-dimensional static visual displays. We ignore auditory or tactile temporal stimuli, where the mathematics take a different form, or three-dimensional and/or temporally varying visual stimuli, which add either another spatial or another temporal dimension to the two-dimensional complexity we are trying understand.

Our evidence shows that the JPEG file size of a visual display is correlated to judgments of visual complexity and to performance on visual display tasks. We found that in search tasks, performance was monotonically related to JPEG file size, but that in same-different tasks, the relationship was non - monotonic. We believe that our measure of complexity is analogous to the measure of algorithmic complexity developed by Solomonoff. He defines the algorithmic complexity of a message as the size, p, of the file necessary to generate that message (in our case, a two-dimensional display) if the file is used as input to a Turing machine and the output of the Turing machine is the generated message. The larger the input file size p, the greater the algorithmic complexity of the output message . In our case, the file whose size we use as an index of complexity is produced by the JPEG image compression algorithm. The JPEG algorithm is like the inverse of a Turning machine. It produces a generating file from a complete (bitmap) message by simplifying the message down to some minimalist criterion. In our case, we use lossless compression to produce this "minimalist" file. Where a Turning machine generates a complete message from a minimalist file, the JPEG algorithm generates a minimalist file from a complete message. In both cases, the file size of the minimalist file is the measure of the complexity of the complete image.

Why does this work for visual images? The JPEG algorithm computes the two-dimensional Fourier transform of separate 8x8 pixel segments of the entire image and stores the coefficients in a file which is then simplified further by a linear compression algorithm (Salomon, 1998). In images that are "simple" – that have large areas of the same colour, many of the higher frequency coefficients will be identically zero and can be compressed easily by a linear compression algorithm. Visual system neurons are activated at all levels of the visual system by input contours. A contour is a contrast border where two different luminances ("colours") meet. Such a contrast border adds high-frequency coefficients to the JPEG transformation, and also adds activity in the visual system. The JPEG file size measure is a direct measure of how much contour variety there will be in the visual display image, and therefore an indirect measure of the amount of stimulation that neurons in the visual system receive from the display.

However, our data show that ZIP file size is highly correlated to JPEG file size, and so the compressed file size generated by either method would be an equally good predictor of LDT The invariance theorem of complexity theory states that the shortest description of any object is invariant (up to a constant) between different universal languages (Chater, 1996), and JPEG and ZIP compressed file sizes are approximations of different "universal languages." We know that JPEG and ZIP file sizes are equally good predictors of rated complexity (Donderi and McFadden, 2003).

JPEG Complexity, Algorithmic Complexity and Random Complexity

We have shown that the subjective complexity of certain types of images is related to the file size required to code that image using the lossless JPEG compression algorithm. We also said that the JPEG algorithm acts like an inverse Turing machine because it generates a program (the JPEG file) from an image, while the Turing machine generates

an image from a program (the shortest program, p). In both cases the program length measures the complexity of the image.

A random string of symbols cannot be generated by a Turing machine using a program that is shorter than the string itself, nor can a random string of pixels be compressed by the JPEG algorithm into an image file that is smaller than the random image. It follows that a random string is maximally complex according to algorithmic complexity, and a random image is maximally complex according to JPEG complexity.

A non-random structure is what enables a program of length p (less than the length of the symbol string s) to produce the string s when fed to a Turing machine, and a non-random structure is what enables the JPEG algorithm to generate a JPEG file that is smaller than the original image file.

Evidence suggests that human complexity judgment is inconsistent with both algorithmic and JPEG complexity. Random images are judged to be less complex than images that are structured in complex, non-random ways. Or, as Wolfram (2002) suggested, the perceived complexity of an image is a function of the most complex structure in the image that humans can detect. As the spatial frequency of the non-random structure increases and the absolute size of the structural elements decrease the structural components become less discriminable because human visual sensitivity is reduced at high spatial frequencies. We cannot detect the structure in a random image because it is not there, and we cannot detect the structure in a very high-frequency non – random image because it cannot be seen.

This suggests that the subjective complexity of an image will be at a maximum somewhere between simple order and complete randomness, and that complete order and complete randomness are not that perceptually different. In the extreme, an entirely black computer screen would be judged not much less complex than a screen generated by a random collection of red, green and blue pixels. From this it follows that subjective complexity should be an inverted U-function of algorithmic complexity and JPEG complexity.

Unanswered questions

Our research leaves unanswered some important questions about the usefulness of JPEG file size as a measure of visual complexity.

1. What is the performance function relating the complexity of a search target and the complexity of the display in which the target is embedded ?

Although the targets in our search experiments were simple, search targets are not always simple. A search target could be a complex image embedded in a complex display background. There may be a non-trivial relationship between the complexity of the target taken by itself, the complexity of the background display taken by itself, and performance measures on a search task involving the target and the display. Our results in the fast

same-different task suggest this. One way to describe this task is as "searching for the chart in the radar display." Decision accuracy in this experiment was best at intermediate levels of chart (search target) and radar (background display) complexity. This same thing might happen when the search target and the display occupy the same space at the same time.

2. Can a display become so "complex" that it appears "simple?"

This apparently contradictory idea is based on Steven Wolfram's discussion of the complexity of one-dimensional cellular automata. He pointed out that the more complex the generating conditions for a cellular automaton, the less easy it was to perceive the pattern in the automaton. This observation matches the expected relationship between a completely random display and the complexity measure that would be generated for it either by a Turing machine or by the JPEG algorithm. Since there is no redundancy in a random image, there is no way to compress it. The generating file, p, that a Turing machine requires to produce that image will be as large as the image itself (Wolfram, 2002). The JPEG compressed file, under lossless compression, will be as large as the randomly generated display file itself. In both cases, the measure of complexity will be as large as it can be, given the display file size. Yet perceptually, a completely random display will be perceived as not much more complex than a completely uniform display.

It is probably premature to interpret the inverted – U relationship between LJPG and LDT in Experiment 4, and between LJPG and LDT and complexity in experiment 5, as confirmation of the idea that as LJPG (file complexity) increases beyond a certain point, the result is perceptually "simpler." That is because the effect is most likely due to the unique JPEG processing of certain curved symbols, and not to a general principle observed in these experiments. Also, the chart symbol displays had LJPGs that were much smaller than those of the electronic chart or aerial photo displays, and so it is clear that the effects are probably experiment – specific, rather than general.

We do not believe that either our Tullis displays or our marine chart, radar and overlay displays reached a level of complexity where the perceptual patterns in the displays began to disappear in the complexity. We do not know whether, if we study displays with even higher levels of JPEG file complexity, we will begin to see a reduction of perceived complexity and a concomitant change in performance. In short, we do not know where to locate the inflection point that is implied by the Wolfram's observation about the perceptual simplicity of even more complex visual displays.

3. What is the balance between structural and semantic complexity and task difficulty for symbolic displays?

In experiment 4, which imposed a rigorous attention demand on the observer, both structural and semantic information contributed to LDT. In Experiment 5, which was less demanding, LDT was a function of semantic information only, but subjective complexity, rated by independent observers, was determined by both semantic and structural information. The chart symbol displays represent a realistic class of displays, and it will

be useful to better understand the relationship between task demands and the measurement of structural and semantic information in such displays.

REFERENCES

Arthus-Bertrand, Y. (1999). *Earth from above.* New York: Harry N. Abrams

Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods and results.* New York: Holt Rinehart and Winston.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, 103(3), 566 – 581.

Donderi, D.C. & McFadden, S. (2003) A compressed file size measure of visual display complexity. Unpublished MS.

Donderi, D.C. (2000). A practical metric for the complexity of graphic displays. *Experimental Psychology Society-BBCS Joint Meeting, Cambridge, UK July 2000* (abstract).

Donderi, D.C. (1999). *An experimental study of marine radar and electronic chart integration.* DCIEM-CR-1999-055, Report to the Defence and Civil Institute of Environmental Medicine, Toronto, Ontario.

Donderi, D.C. & McFadden, S. (1999). Performance measurement of combined versus separate radar and electronic chart displays. *19th International Cartographic Conference, International Cartographic Association*, Ottawa

Donderi, D.C. (2001). *The use of electronic charts and radar in marine navigation.* DCIEM CR 2001- 03, Report to the Defence and Civil Institute of Environmental Medicine, Toronto, Ontario

Leeuwenberg, E.L.L. (1969). Quantitative specification of information in sequential patterns. *Psychological Review*, 76(2), 216 – 220.

Leeuwenberg, E.L.L. (1968). *Structural information of visual patterns.* The Hague: Mouton.

McFadden, S., Bauer, B., & McManus, K. (1997). Measures of the Discriminability of Symbol Shapes. DCIEM No. 97-R-61, North York, Defence and Civil Institute of Environmental Medicine.

Quastler, H., ed. (1954). *Information theory in psychology: problems and methods.* Glencoe, IL, Free Press.

Salomon, D. (1998). *Data compression: the complete reference.* New York: Springer.

Solomonoff, R.J. (1986). The application of algorithmic probability to problems in artificial intelligence. In Kanal, L. N. & Lemmer, J. F. (eds.). *Uncertainty in artificial intelligence*. North-Holland: Elsevier.

Stevens, S.S. (1975). *Psychophysics: Introduction to its perceptual, neural and social prospects*. New York: Wiley

Tullis, T.S. (1983). The formatting of alphanumeric displays: A review and analysis. *Human Factors*, 25, 657-682.

Wolfram, S. (2002). *A new kind of science*. Champaign, IL: Wolfram Research

DOCUMENT CONTROL DATA SHEET

1a. PERFORMING AGENCY
Human Factors North Inc., 118 Baldwin Street, Toronto, Ontario, M5T 1L6

2. SECURITY CLASSIFICATION
UNCLASSIFIED
Unlimited distribution -

1b. PUBLISHING AGENCY
DRDC Toronto

3. TITLE
(U) A complexity measure for electronic displays: Final report on experiments

4. AUTHORS
Don C. Donderi

5. DATE OF PUBLICATION
March 4 , 2003

6. NO. OF PAGES
68

7. DESCRIPTIVE NOTES

8. SPONSORING/MONITORING/CONTRACTING/TASKING AGENCY
Sponsoring Agency:
Monitoring Agency:
Contracting Agency : DRDC Toronto
Tasking Agency:

9. ORIGINATORS DOCUMENT NO.
Contract Report CR 2003-046

10. CONTRACT GRANT AND/OR PROJECT NO.
W7711-017758/001/TOR

11. OTHER DOCUMENT NOS.

12. DOCUMENT RELEASABILITY
Unlimited distribution

13. DOCUMENT ANNOUNCEMENT
Unlimited announcement

14. ABSTRACT

(U) Five experiments studied the relationship between compressed file size and search performance for visual displays. Experiments 1 and 2 were visual search tasks where observers searched displays to find either a triangle or a parallelogram randomly positioned on the display. In experiment 1 the displays were marine electronic charts, radars, or chart-radar overlays. In experiment 2 they were colour photographs of the earth. In both experiments, search decision time, rated complexity, and proportion of decision errors were monotonically related to the log JPEG compressed file size (LJPG) of the displays. Experiments 3 looked at two different same-different decision tasks in which a marine chart display was followed by a marine radar display. The task was to decide whether they represented the same or a different area. The ad lib condition gave unlimited viewing time while the fast condition gave a short viewing time for each display. Chart viewing time in the ad lib condition was a positive increasing function of chart LJPG. In the fast condition, decision accuracy depended on chart and radar LJPG and the subjective similarity between chart and radar pairs. Displays in the search tasks of experiments 4 and 5 presented groups of chart symbols containing either triangles or parallelograms as targets and other symbols as background. Decision time in both experiments was non-linear in LJPG, but linear in structural and semantic information content (Experiment 4) or semantic information content (experiment 5). The uses and limitations of compressed file size as a measure of display complexity are discussed.

(U) Cinq expériences ont été menées en vue d'étudier la relation entre la taille des fichiers compressés et la performance de recherche dans des affichages. Dans les expériences 1 et 2, les observateurs devaient exécuter des tâches de recherche visuelle qui consistaient à examiner des affichages pour trouver un triangle ou un parallélogramme positionné de façon aléatoire dans l'affichage. Les affichages utilisés dans l'expérience 1 étaient composés de cartes électroniques marines, de cartes radar ou de recouvrements carte-radar. Dans l'expérience 2, des photographies couleur de la Terre ont été utilisées. Dans les deux expériences, la durée de la recherche, la cote de complexité et la proportion des erreurs de décision étaient monotoniquement reliées à la taille des fichiers JPEG compressés (TFJC) consignée pour les affichages. L'expérience 3 comprenait deux tâches différentes de distinction qui consistaient à déterminer si une carte marine et une carte radar marine affichées consécutivement représentaient le même emplacement ou un emplacement différent. Dans l'expérience réalisée en condition libre, la durée d'examen des affichages n'étaient pas limitée, tandis que dans l'expérience réalisée en condition rapide, les observateurs ne disposaient que d'un court laps de temps pour examiner chaque affichage. En condition libre, la durée d'examen de l'affichage était une fonction positive croissante de la TFJC de la carte. En condition rapide, l'exactitude de la décision dépendait de la TFJC de la carte marine et de la carte radar ainsi que de la similarité subjective entre les paires carte marine-carte radar. Pour les tâches de recherche des expériences 4 et 5, les affichages montraient des groupes de symboles de carte dans lesquels des triangles ou des parallélogrammes étaient utilisés comme symboles cibles et d'autres symboles étaient utilisés comme symboles de fond. Le délai de décision (DD) obtenu dans les deux expériences n'était pas une fonction linéaire de la TFJC, mais une fonction linéaire du contenu de l'information structurelle et de l'information sémantique (expérience 4) ou du contenu de l'information sémantique (expérience 5). Le présent document traite de l'utilisation et des limites de la taille des fichiers compressés comme mesure de complexité des affichages.

15. KEYWORDS, DESCRIPTORS or IDENTIFIERS

(U) complexity; visual displays; electronic displays; human performance; visual search; JPEG compression; structural complexity; subjective complexity; electronic charts